A STUDY OF THE SILICON IMPATT DIODE

by

William Elder Wilson Tait, M.Sc.

Thesis presented for the Degree of
Doctor of Philosophy of the University
of Edinburgh in the Faculty of Science.

July 1972

TABLE OF CONTENTS

3.3.2/

SUMMARY

The purpose of this work is to develop a low cost technology which will permit the fabrication of abrupt junction silicon IMPATT diodes with microwave generation capabilities at X-band frequencies.

As an introduction, the recent miniaturisation of microwave systems is reviewed briefly to indicate the importance of a solid state source in the progression to microelectronic media. Emphasis is placed on the potential role of the IMPATT diode, as various structural configurations are discussed historically and in relation to the requirements of this research.

A theoretical analysis of the p^+nn^+ diode demonstrates the essential dynamic negative resistance property, and design criteria involved in silicon impurity profiling are explained during the determination of an ideal structure designed for IMPATT operation at 9.5 GHz.

Technologies available for silicon impurity doping are described, and inherent limitations in each process related to the practical consideration of abrupt junction approximation. Other experimental factors of equipment cost and accessibility are explained, and a technique developed for accurate p-type aluminium alloy doping of epitaxially grown n on n^+ silicon substrate material.

Problems associated with ensuring adequate device

dissipation are mentioned as component fabrication is undertaken. Processed silicon slices are maintained in an integral form during thermocompression attachment to metallic heat sinks, and various inexpensive methods of device definition and separation are investigated in the production of complete diodes.

A test facility is established for microwave detection, and results are presented for pulsed operation to device limits. Restrictions in output and component yield are intrepreted in terms of fabrication procedure: some remedial experiments are performed, and recommendations made for future improvement.

CHAPTER 1

INTRODUCTION

1.1 THE CONCEPT OF MICROWAVE INTEGRATED CIRCUITRY

In the past, equipment operating at microwave frequencies has been characterised by massive assemblies of precision machined components interconnected by waveguide or coaxial cable. Designs have called for tight tolerances to allow suitable electrical performance and heavy walled structures were necessary to withstand environmental stresses. Furthermore, complex field patterns arising from irregular geometries and machining inaccuracies meant that electrical properties had to be derived empirically rather than theoretically, and equipment development times could be protracted as a result of this unsatisfactory procedure.

While this situation perpetuated, lower frequency electronics was being revolutionised by emergent microelectronic technologies providing semiconductor integrated circuits consisting of active and passive components prepared from the same semiconductor crystal; thick and thin film processes enabling microminiature passive elements on inert substrates; and a hybrid construction where the expedient of combining the forementioned disciplines allowed attributes of each process to be incorporated in the realisation of a circuit function. Such techniques brought benefits of reduced cost, increased durability

and reliability, precision tolerance control, and lower size and weight which have increased modular complexity and versatility in modern equipment design.

Applications for integration of microwave systems can be found in both commercial and military equipment, where reduced design and fabrication costs would expand utilisation to areas presently precluded by prohibitive expense. The commercial sector would benefit from microwave links, competitive in cost with current V.H.F. equipment while providing increased channel capacity and interference rejection through directivity; whereas military equipment would experience reduced size and weight, permitting redundancy as a system reliability measure and increased ruggedness which must accompany such compact construction.

Although the field of microwave engineering has been the greatest potential beneficiary of these techniques, it has seldom exploited their advantages to the present day. While semiconductor integrated circuitry was a natural progression from established discrete semiconductor devices, little comparable background existed in the realm of microwave technology because of poor transistor performance at such high frequencies, and the problem of replacing microwave tube power sources (klystrons and magnetrons) with solid state equivalents. Nevertheless, the concept of microwave integrated circuitry has been a subject of considerable engineering interest during recent years, and resultant progress will now be discussed.

Several solid state microwave components have already reached sufficient sophistication for this purpose; indeed the point contact diode, being one of the earliest semiconductor devices investigated, has been widely used as a detector of amplitude modulated radio frequency signals and as a mixer in microwave receiver systems. More recent work on the Schottky barrier diode has revealed many superior properties; while other functions such as switching, limiting, phase-shifting, and amplification can be readily achieved using p-i-n diodes, tunnel diodes, varactor diodes, and to a lesser extent transistors.

Despite the existence of these devices, it has proved difficult to generate modest power levels at microwave frequencies, and devious indirect means have been sought to overcome this problem.

One such solution consists of a crystal controlled oscillator, transistor amplifier parametric multipliers, and varactor multipliers coupled to raise the oscillation frequency from megahertz to gigahertz ranges. This combination can provide useful output of the order of watts at X-band frequencies providing the transition from transistor to varactor stages is made within the limits of high power high frequency transistor operation. Extreme care must be exercised in obtaining good impedance matching between stages however, and filtering or appropriate isolation must be included to suppress all unwanted harmonics.

A simpler technique involves the tunnel diode permanently

biased in the negative region of the current-voltage characteristic to provide high frequency oscillation. Although this generator has low noise properties, the power output capability is inherently lower than obtainable by varactor harmonic generation, typically below 100 mW at X-band.

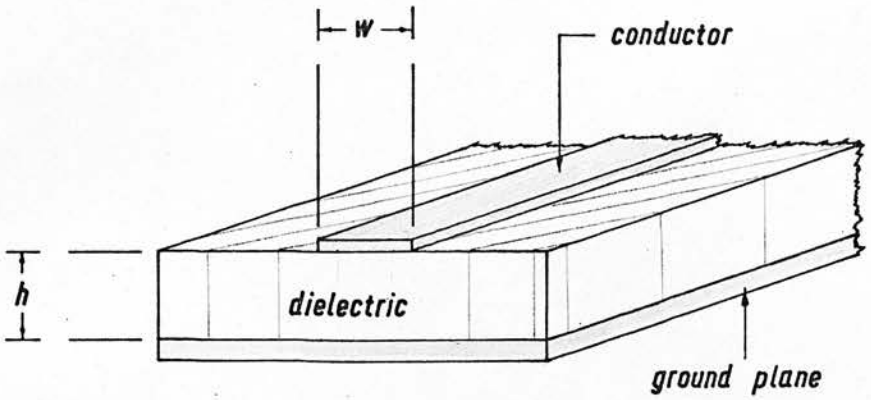
These approaches have been the established sources of solid state microwave power, and the absence of higher generation levels, both C.W. and pulsed, has limited the range of application for integrated microwave systems.

During recent years however, the development of two new semiconductor devices - the Gunn diode and the avalanche diode have improved this situation, and generation levels several orders of magnitude higher can be obtained with a useful device yield.

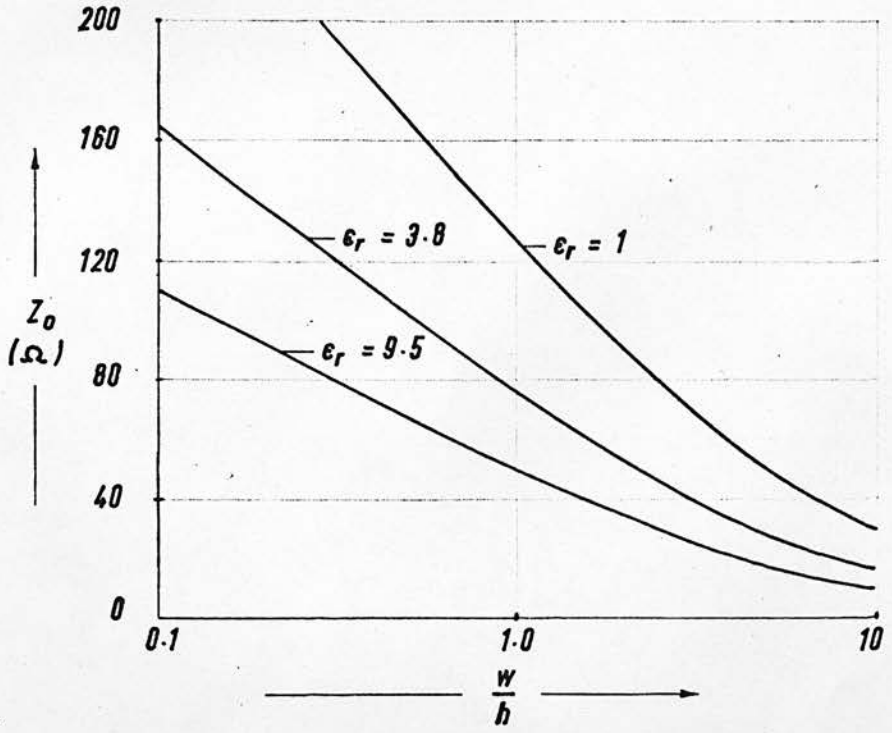
The Gunn diode relies on conduction band properties of semiconductors formed from compounds in groups *III - V* and *II - VI* of the Periodic table, to enable high frequency oscillation when a field is placed across the sample in excess of a threshold value. This oscillation, associated with a moving layer of charge (the domain mode), has a period approximately equal to the transit time of carriers between device contacts. Alternative excitation involving the suppression of full domain formation by the application of an appropriate r.f. voltage waveform does not depend on carrier transit time and this limited space-charge accumulation (L.S.A.) mode has been used to generate larger power levels at higher frequencies than can be achieved under Gunn operation.

While Gunn diodes are produced almost exclusively from gallium arsenide, a semiconductor less technologically advanced than silicon and restricted in its ultimate application, considerable interest has been shown in the avalanche diode which can be fabricated using well developed manufacturing techniques to provide a microwave power source with greater potential for integration. This diode employs the cumulative delays arising from carrier population build up and the transit time of carrier movement between device contacts under avalanche reverse bias to provide a phase delay of current in an external circuit which can constitute a negative resistance and cause spontaneous oscillation. Such an operational mode is commonly termed IMPATT (an acronym for IMPact Avalanche Transit Time) and, at the time of commencing the present study, commercially available IMPATT diodes were capable of delivering several hundred milliwatts at X-band when placed in the correct environment. Within research laboratories considerably higher outputs have been reported, and the discovery of a more efficient, lower frequency 'anomolous mode' has increased interest in diodes of this structural type.

In the field of passive components, the use of transmission lines as lumped resistor, capacitor, and inductor counterparts places certain constraints on the nature of integration. A configuration which achieves compatability is microstrip, shown in the form of Fig. 1.1a to consist of a conductor and electrical ground plane separated by a dieletric region which contains all but the fringing fields and controls microwave propagation



(a) Microstrip Schematic



(b) Graph of Z_o against $\frac{w}{h}$ as a function of ε_r

Fig. 1.1

velocity. In fabrication, conductor patterns may be established from a uniform metallic overlayer using a precision photolithographic technique developed as part of the conventional integrated circuit process, and reproducibility within realistic limits of $\pm$ 2 microns (μm) is certainly adequate.

An appreciation of the difficulties in microwave integration must involve an understanding of electrical parameters associated with microstrip and the imposition these place on the choice of constituent materials. Comprehensive literature has been provided on this topic, notably by Wheeler [1] and Yamashita et al. [2] who have researched the theoretical aspects of microstrip design, and by Presser [3] who has summarised practical design rules for construction purposes. It is sufficient here to quote approximate relationships between (i) Z_o , the line characteristic impedance, and ε_r , the relative dielectric constant of the intermediate layer, in terms of assembly geometry:

$$Z_o = 377 \frac{h}{W} \sqrt{\frac{1}{\varepsilon_r}} \quad \text{ohms} \qquad 1.1.1$$

where h is the thickness of the intermediate layer and W is the conductor width;

(ii) α_c, conductor attenuation as a function of conductor width (W), characteristic impedance (Z_o), and microwave frequency (f):

$$\alpha_c = 2{\cdot}6 \times 10^{-7} \frac{\sqrt{f}}{W Z_o} \quad \text{nepers/metre} \qquad 1.1.2$$

(iii) λ_a , the active wavelength, in terms of dielectric constant (ε_r) and frequency (f)

$$\lambda_a = 3 \times 10^8 \frac{1}{f\sqrt{\varepsilon_r}} \quad \text{metres} \qquad 1.1.3$$

Equation 1.1.1 forms a basis for microstrip design and can best be expressed graphically as in Fig. 1.1b. Within this framework however, geometric restrictions are imposed for various reasons, the more important of which will now be mentioned.

A limiting feature in miniaturisation arises from increasing line loss associated with reduction in conductor width according to equation 1.1.2. Design of useful transmission lines with suitable characteristic impedances is therefore restricted to a minimum conductor line width of approximately 50 μm and dielectric thickness greater than 250 μm.

From equation 1.1.3 it will be realised that line length in microstrip is an inverse function of the intermediate layer dielectric constant. Consequently, high dielectric materials must be selected to minimise substrate area requirements: $\varepsilon_r > 9$ is normal.

In addition to these constraints demonstrated mathematically, it is necessary to ensure low microwave dielectric loss , high thermal conductivity, and integration compatibility of any substrate material. From this background, consideration can be given to various assembly techniques, and problems

associated with each alternative will be clearly understood.

In many respects, the constructional options in microwave integration are similar to those of conventional integrated circuits, and the choice lies between a monolithic technique where the passive circuit and active devices may be batch fabricated simultaneously without the need for bonding connections between elements, and a hybrid assembly where passive circuitry is batch fabricated on a substrate and active devices are subsequently bonded in position.

Monolithic fabrication has already been attempted in several ways. Use of high resistivity silicon as the dielectric into which active devices may be formed, offers the opportunity of monolithic construction where passive components are subsequently deposited on the silicon surface. It is exceedingly difficult to achieve low loss using silicon as a dielectric since resistivity decreases with device processing and temperature. This thermal problem is further aggravated by formation of active devices in the upper region of the silicon, since device separation by almost the entire high resistivity substrate thickness inevitably gives rise to unsatisfactory heat sinking. In general, an all silicon technology suffers from a lack of versatility where other dielectric and semiconductor materials might be required to ensure adequate circuit performance.

An alternative monolithic construction makes use of low resistivity silicon as a ground plane and substrate for epitaxial

silicon growth into which active devices can be formed. Thermal oxidation of the surface, vacuum deposition of a suitable dielectric, etching of contact windows, and final thin film metallisation completes a circuit free from problems of heat sinking and lossy dielectrics. Nevertheless, the use of other semiconducting materials is still precluded.

Selection of a dielectric which has a crystallographic structure capable of supporting epitaxial growth of several semiconductors will allow this versatility in active device fabrication. The 'silicon on sapphire' technique permits the production of very small geometry devices in extremely thin epitaxial layers which can then have passive elements superimposed by vacuum deposition to complete the circuit. In this application, the process classification is misleading since its attribute lies in the ability to create not only localised deposits of silicon, but also of other semiconductors such as gallium arsenide. Although the construction combines low dielectric loss and multiple semiconductor integration, it does re-introduce a problem of inadequate heat sinking since devices are separated from a ground plane by the entire sapphire thickness.

The alternative integration philosophy of hybrid construction is characterised by discrete devices, either wire- or die-bonded to a passive circuit which is a product of thin or thick film technologies.

A thin film process involves vacuum deposition of conductive,

resistive, and dielectric layers on a low loss substrate which is highly polished to allow accurate photoetching and prevent rough surface losses. While quartz has been the most suitable material for this work, its relatively low dielectric constant has restricted useful designs to frequencies above 5 GHz, and only recently have improvements in the surface finish available on higher dielectric alumina substrates increased the potential range of operation.

In thick film processing, conductive, resistive and dielectric 'inks' are screen printed and fired on a ceramic substrate selected from a wide range of suitable materials. Although lithographic accuracy does not permit the precise geometry obtainable with thin film construction, the variety of material dielectric constants does assist in circuit miniaturisation, and impedance levels for optimum dimensions tend to be lower, enabling larger area, lower impedance semiconductor devices capable of handling more power. Because of the relatively rough surface profile of ceramics and fired layers in thick film production, the technique is restricted to X-band (8-12 GHz) and lower frequencies.

This superficial discussion has demonstrated that no single technology combines all the features which are desirable in a microwave integrated circuit. The situation, although partly analogous to the dilemma of conventional integration, remains weighted against a monolithic construction by the present low yield

experienced in active device formation and problems of power dissipation accompanying inefficient operation at microwave frequencies. It is the existence of such difficulties which has increased interest in a hybrid construction allowing evaluation and selection of low yield components, the use of various dielectrics and semiconductors; and permitting heat sinking precautions where necessary.

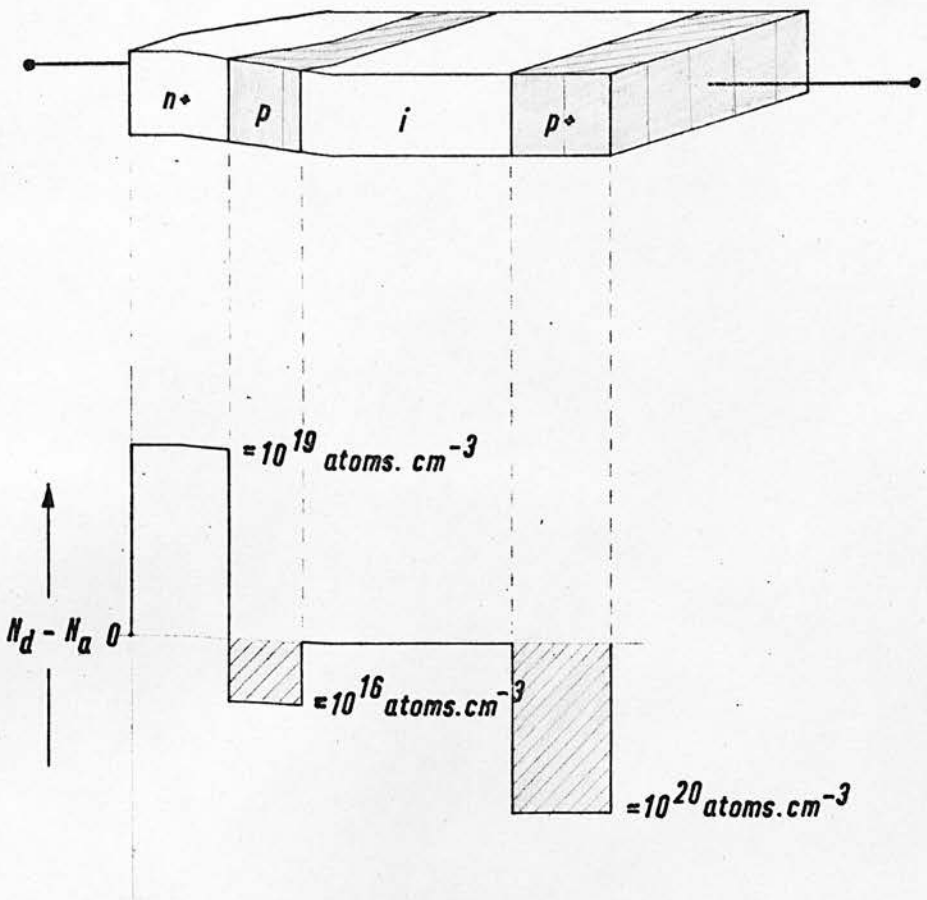
Future development must rely on improved manufacture of discrete solid state microwave devices before predictable components can be properly assessed in determining a suitable integration philosophy. This thesis is devoted to the study of one such device, the IMPATT diode, which appears to have the greatest application potential of all ~~mircrowave~~ microwave active devices currently within research and development laboratories. Where silicon is used as a fabrication medium, component requirement can be clearly defined in terms of high yield semiconductor processing which can later be applied to monolithic integration of low power circuits, and a compatible assembly technique providing high power discrete devices for hybrid applications. Most attempts to achieve these properties have been associated with existing silicon I.C. technology: this work examines alternative processes in an endeavour to contribute a further fabrication schedule which may provide higher yield devices exhibiting improved heat sinking properties.

1.2 THE HISTORICAL DEVELOPMENT OF IMPATT DIODES

In 1954, Shockley [4] introduced the concept of impulsive

impedance of a two terminal semiconductor device operating under transient conditions to demonstrate analytically that transit time carrier movement in a suitable diode structure could give rise to negative resistance properties.

Four years later, Read [5] developed this theory in his study of a diode structure designed to produce oscillations in the 1-50 GHz frequency range when mounted in a suitable microwave cavity. The proposed $n^+ pip^+$ configuration, shown in Fig. 1.2.1, was to be reverse biased causing carrier depletion in a high resistance region which was bounded by very low resistance end regions. In such a situation the electric field maximum occurs at the edge of the depletion region and may be arranged large enough to initiate electron-hole pair generation from the injected carriers by the well understood secondary emission or avalanche process. The applied voltage is always above a 'punch through' level so that the space charge region extends from the n^+p junction to the ip^+ junction. While electrons generated within the 'avalanche zone' travel directly to the n^+ region, the holes move across the remainder of the space charge region at saturated scattering limited velocity because of the electric field magnitude. Read recognised a build up time for the avalanche process and a transit time for holes travelling across the depletion region, which by design, could provide a cumulative delay in external circuit current of one half cycle relative to an initiating a.c. voltage. In this way, power could be delivered to the a.c. signal, and providing the device was mounted in an



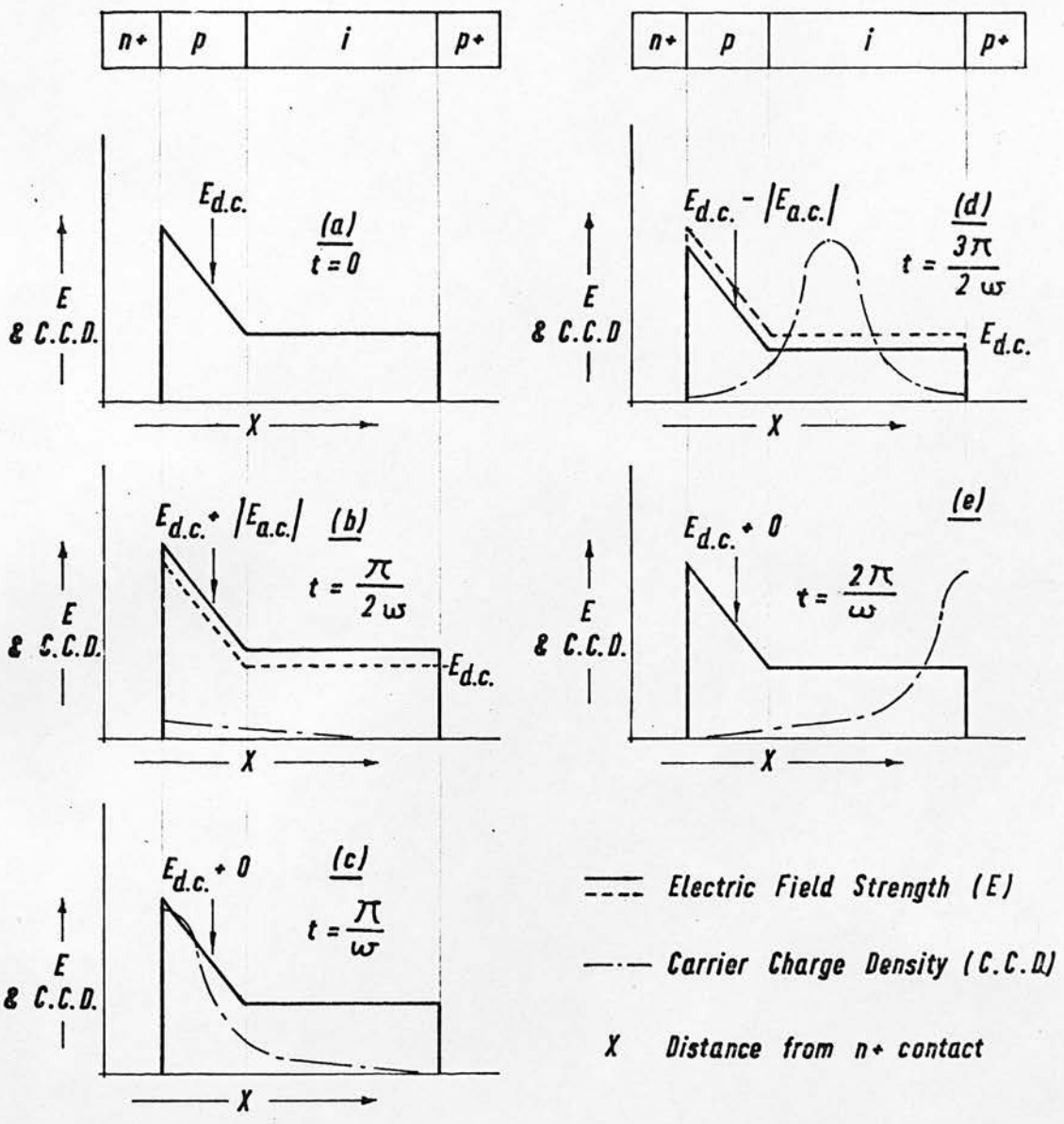
Idealised Read Diode

Fig. 1.2.1

inductive microwave cavity tuned to the capacity of the diode, an oscillation would occur.

The mechanism of energy transfer may most easily be understood by reference to Fig. 1.2.2. The simplified graph of electric field within the diode shown in (a) depicts the situation occurring when an n^+pip^+ diode is reverse biased to a level just below that required for avalanche. With the exception of a reverse saturation current, no diode current will flow. The effect of superimposing an a.c. voltage ($V_{a.c.}$) upon the existing d.c. voltage ($V_{d.c.}$) will now be considered, where $|V_{a.c.}|$ is sufficiently large to allow avalanche carrier multiplication during the positive half cycle.

At the instant t = o, a small number of mobile carriers are present at the n^+p junction resulting from reverse saturation current only. As the a.c. component increases ($t=o+\delta t$), carrier multiplication commences in the high field region, electrons travel to the n^+ region and holes enter the drift region. Fig. 1.2.2 (b) indicates the situation present at $t = \frac{\pi}{2\omega}$. Since the number of carriers initially present for multiplication is small, the exponential growth rate expected has not contributed significantly to the carrier charge density during a quarter cycle. This is very fortuitous in that carriers traversing the drift region will have work done on them by the a.c. field. Such an undesirable situation continues during the next quarter cycle $\frac{\pi}{2\omega} \leq t \leq \frac{\pi}{\omega}$, as carrier growth proceeds and the a.c. field



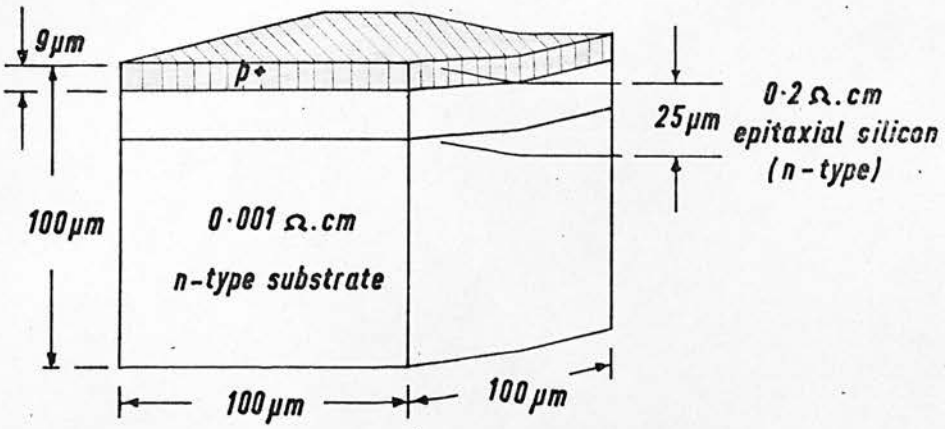
Field and Charge Transport Sequence for a Read Diode

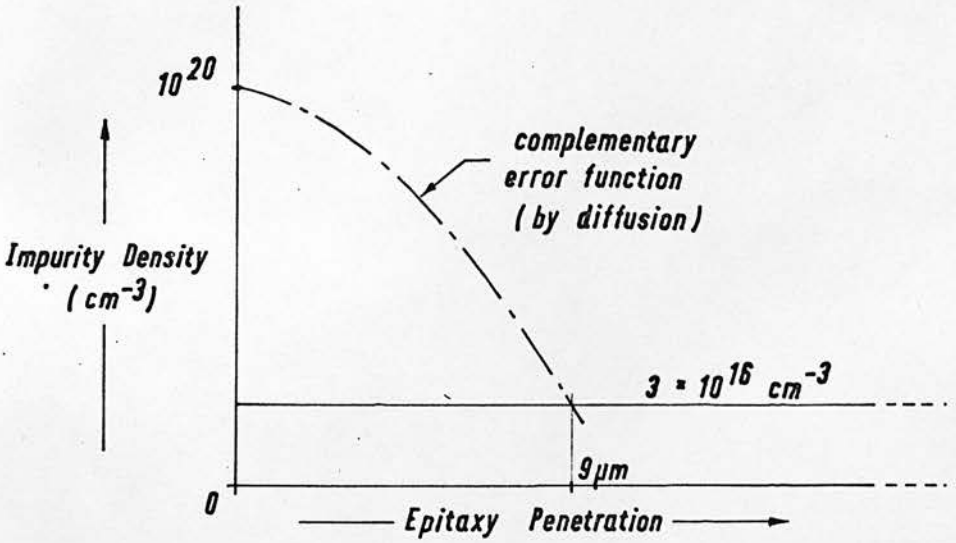
Fig. 1.2.2

continues to assist the drift process. At $t = \frac{\pi}{\omega}$ (Fig. 1.2.2 (c)), a situation exists where the mobile hole density is at a maximum, and the a.c. field changes direction. During the next half a.c. cycle, the incident charge decays and the hole pulse travels across the drift region under the d.c. field, transferring energy to the a.c. component by virtue of its opposite polarity. If the transit time of carriers through the drift region is arranged to be one half the period of the a.c. voltage as indicated in Fig. 1.2.2 (d) and (e), then a net energy transfer from the d.c. to the a.c. field will take place during the complete a.c. cycle, and a self-sustained oscillation may be achieved using appropriate external circuitry.

Despite Read's prediction of high microwave power outputs with high conversion efficiency, attempts to demonstrate the feasibility of this device were thwarted by an inadequate technology giving rise to microplasma effects from field inhomogenities.

In 1965, the first IMPATT operation was reported by Johnston et. al. [6] who observed microwave output from a simple p-n junction diode under pulsed operating conditions. Fig. 1.2.3 shows the structure used to obtain 80 mW power at 12 GHz from 2 μs pulses at 10 kHz p.r.f. For this device, it was suggested that part of the space charge region formed drift space, since the depletion width approximated the drift length predicted by Read for oscillation at this frequency.





Construction of IMPATT Diode (by Johnston et al.)

Fig. 1.2.3

Although later in the same year Lee et al. [7] obtained 100-600 MHz oscillations using a true Read type diode, the earlier experiments had demonstrated that such structural complexity was unnecessary, and that other configurations could possess microwave negative resistance from IMPATT properties. This was confirmed by the small signal theory developed by Misawa [8, 9] which showed that negative resistance could be obtained from junction diodes of any doping profile.

Since earliest reports of IMPATT operation, considerable effort has been devoted to improvements in device fabrication as better semiconductor wafers have become available. Germanium, gallium arsenide, and silicon have all been used as construction media, and a variety of device configurations has emerged in endeavours to optimise performance. 'State of the art' manufacture in silicon, limited mainly by problems of heat sinking, can provide continuous output powers in the region of one watt at X-band frequencies, but processing still gives rise to low device yield and a measure of unpredictability in design performance.

1.3 SELECTION OF AN IMPATT DIODE STRUCTURE

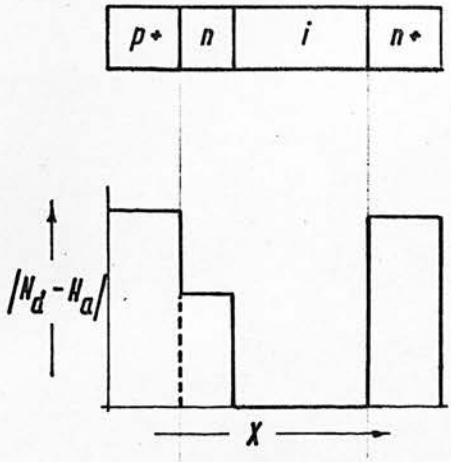
While many diode types can exhibit IMPATT properties, their physical dissimilarities give rise to different design criteria and have severe implications in device processing. A comparison of structures and operating conditions will be made

in order to explain the configuration chosen for this work, and also to introduce some of the design parameters to be considered more fully at a later stage.

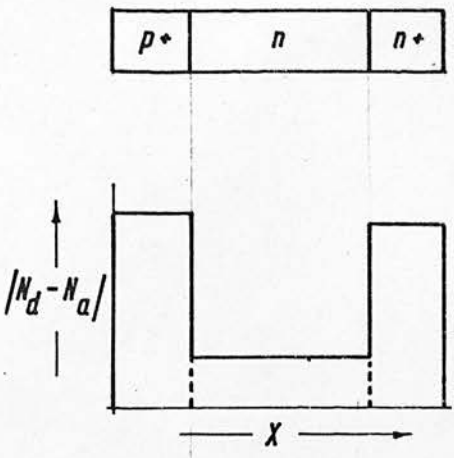
Fig. 1.3.1 (a) shows the impurity doping profile, electric field distribution and ionisation rate within an idealised Read diode. In the avalanche condition, ionisation occurs locally within a small zone at the p^+ n junction, and the electric field is maintained high enough throughout the entire depletion region to ensure saturated carrier velocity transport to the n^+ contact: this minimises transit time and hence recombination. Similar parameters are shown in Fig. 1.3.1 (b) for the p^+ n n^+ structure where distinct avalanche and drift regions also exist. Although the electric field profile is linear and drops below the level required to sustain saturated drift velocity, the low field region is small compared with the total drift length, and the effect on transit time is insignificant. This is also true of the linearity graded p-n junction diode (c), in which ionisation occurs in the centre of the depletion region. It is more diffuse than in any of the abrupt junction diode types, and the lack of symmetry in profile results from the difference in ionisation rate between holes and electrons in silicon. The pin structure (d) possesses a uniform electric field across the intrinsic region: consequently ionisation occurs throughout the entire depletion zone and carrier drift is constant at the saturated velocity level.

In this work, it was intended that use should be made of

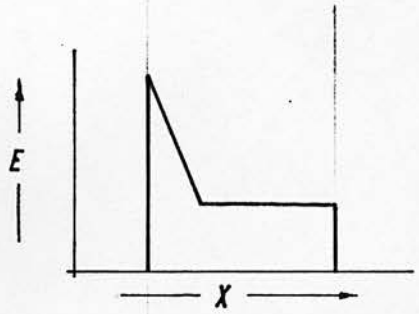
(a) Read Diode

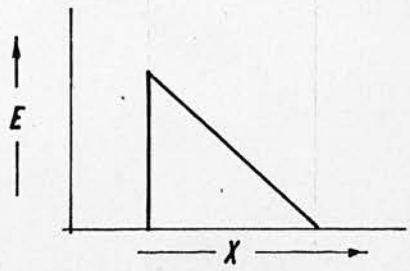


(b) Abrupt p^+nn^+ Diode

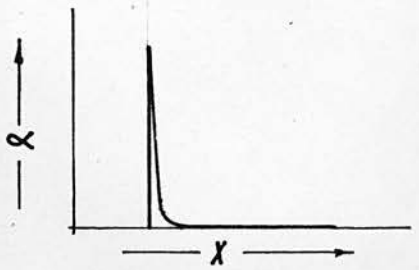


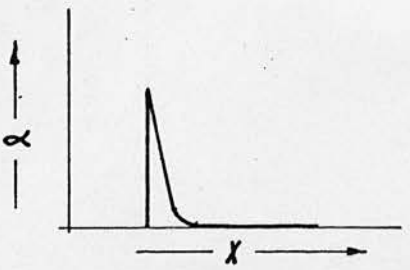
Graph of Impurity Density $\{|N_d - N_a|\}$ against Distance (X) from the p^+ contact





Graph of Electric Field Strength (E) against Distance (X) from the p^+ contact



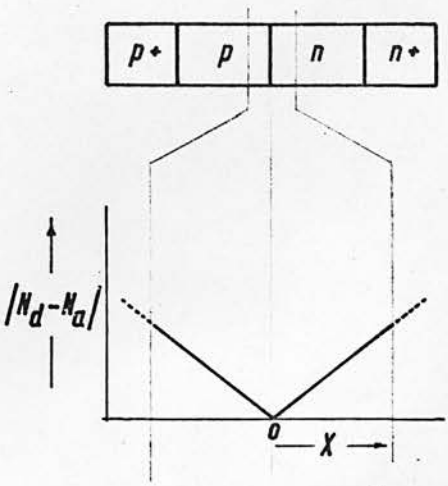


Graph of Ionisation Rate (α) against Distance (X) from the p^+ contact

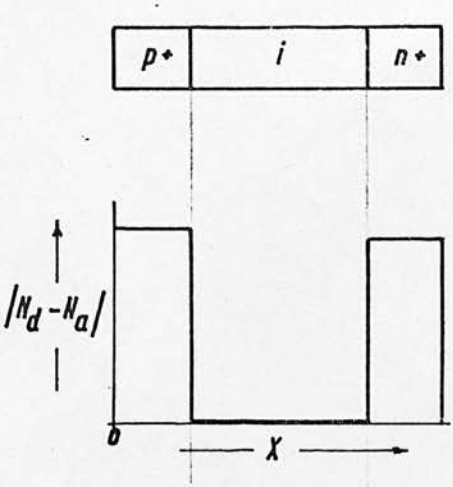
Distribution of Impurity Density, Electric Field Strength, and Ionisation Rate in Read and Abrupt p^+nn^+ Structures

Fig. 1.3.1 (a,b)

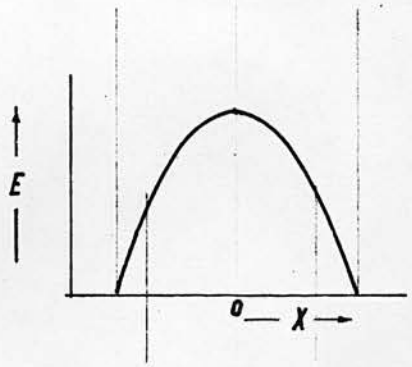
(c) Linearly Graded p-n Junction Diode

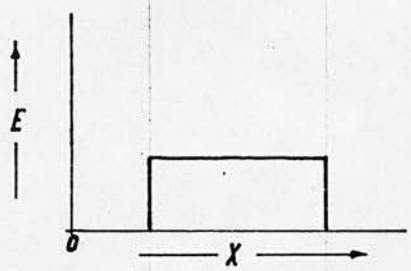


(d) p i n Diode

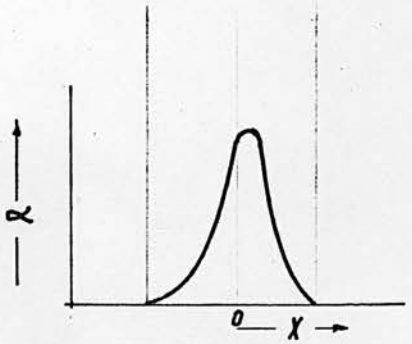


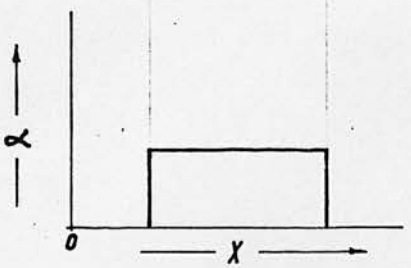
Graph of Impurity Density $\{|N_d - N_a|\}$ against (c) Distance (X) from the p-n junction
(d) Distance (X) from the p^+ contact





Graph of Electric Field Strength (E) against (c) Distance (X) from the p-n junction
(d) Distance (X) from the p^+ contact





Graph of Ionisation Rate (α) against (c) Distance (X) from the p-n junction
(d) Distance (X) from the p^+ contact

Distribution of Impurity Density, Electric Field Strength, and Ionisation Rate in Linearly-Graded and p i n Structures

Fig. 1.3.1 (c, d)

Departmental expertise in abrupt junction formation, and that methods of silicon processing be investigated which hitherto had not been applied to this device. The advantage of structural simplicity is obvious when consideration is give to diagnostic work involved during technology development. Further benefits may be expected in slice yield and minimised device vulnerability during separation and encapsulation.

For these reasons, the complexity of a Read diode was deemed undesirable, and problems associated with production of an intrinsic silicon region for pin formation were thought outwith the scope of this project. In addition, Misawa had shown the improved small signal negative resistance property of a p^+ n junction compared with the pin structure and experimentally demonstrated superior oscillator performance using this device. A p^+n configuration possesses the desired characteristic of simplicity and the modified p^+nn^+ form gives rise to diodes with minimised residual series resistance and improved ohmic terminal properties at the expense of a further processing step. A commitment to this structure seemed most suitable, since it also permits a versatility in process technique which was advantageous for the proposed investigation.

CHAPTER 2

CHARACTERISATION OF THE P^+NN^+ IMPATT DIODE

The demonstration of varying avalanche intensity and location depending on device structure and doping profile illustrates the problem of achieving a unique characterisation and design procedure for IMPATT operation which could be applied to all diode types.

A small signal analysis presented by Read [5] to demonstrate carrier transit time effects from avalanche multiplication in his four layer structure was subsequently modified by Gilden and Hines[10] who examined the IMPATT mechanism in p^+n junctions similar to the type being considered in the present work. Implicit in their analysis is a restricting assumption that carrier multiplication occurs locally. The other extreme case, namely uniform multiplication throughout the space charge region, was examined by Misawa [8, 9], who subsequently combined his theory with the result for avalanche free drift space proposed by Read, to produce a generalised small signal analysis applicable over the entire range of avalanche - drift space interaction.

Large signal diode properties were also considered by Read, however the simplicity of his diode model gave rise to significant inaccuracy in the prediction of high efficiency oscillator application. This over - optimism was revealed by Gummel and Scharfetter [53] during a series of realistic computational analyses incorporating variables which Read had

chosen to ignore in his earlier algebraic treatment.

In view of the silicon doping profiles selected for this study, the theory presented by Gilden and Hines will be used to show the negative resistance property within an abrupt p^+n junction, and the significance of carrier transit time; while large signal properties will be developed from the original Read theory with due regard to corrective findings of Gummel and Scharfetter.

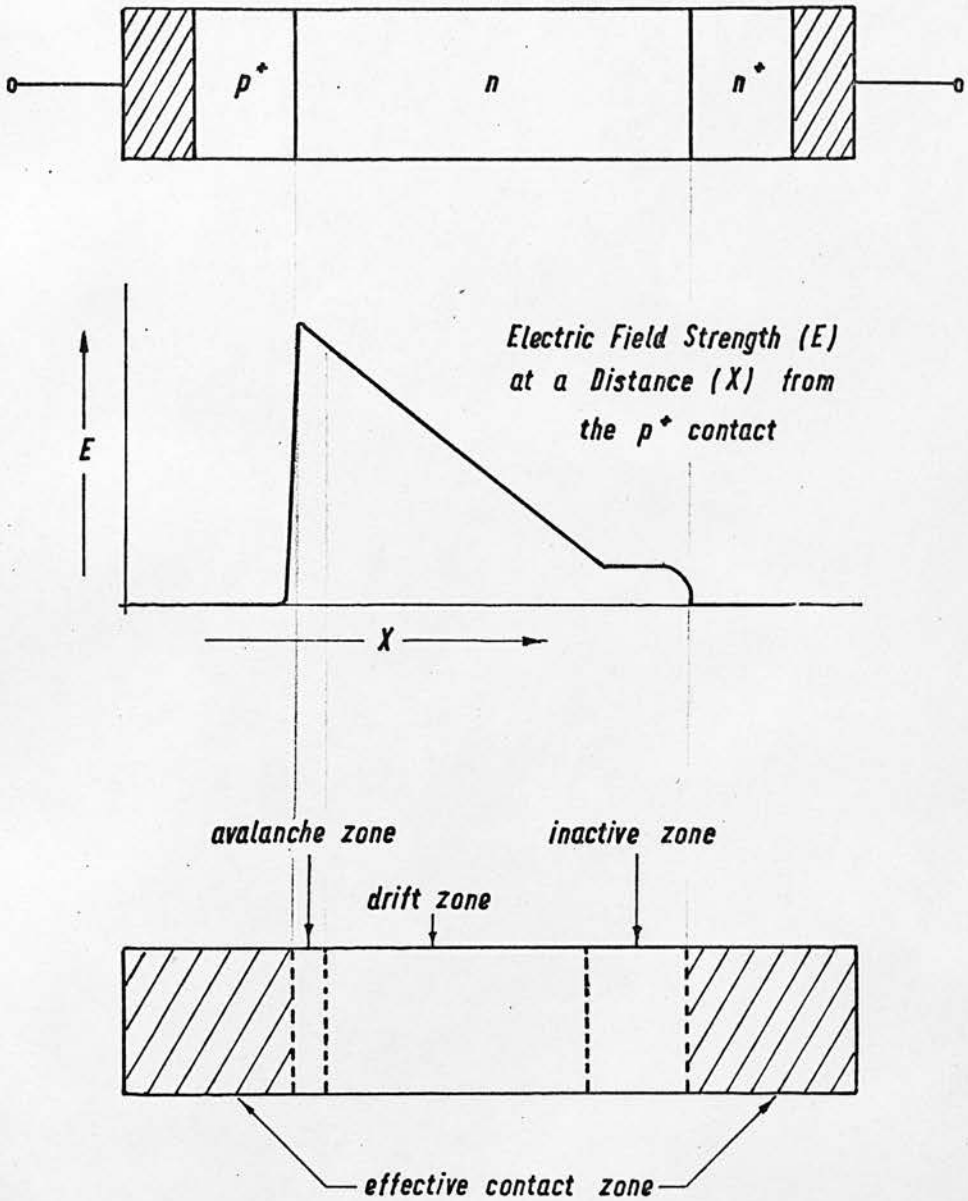
2.1 THE DYNAMIC RESPONSE

(i) A Small Signal Analysis

The model, shown in Fig. 2.1.1 to depict an abrupt junction p^+nn^+ IMPATT diode, has three discrete regions which can be considered as follows:

1. An avalanche zone in which ionisation occurs. It is sufficiently thin that signal transit delay can be ignored.
2. A drift zone which all carriers entering from the avalanche region traverse at scattering limited velocity.
3. An inactive zone which constitutes series resistance associated with any undepleted n-type material and the ohmic contacts.

Each of these regions will be treated independently in determining the response to a small a.c. perturbation, and the derived impedances summed to describe device operation.



Abrupt Junction P+NN Diode Model

Fig. 2.1.1

The dynamics of an avalanching electron-hole plasma may be described using a one-dimensional form of Poisson's equation,

$$\frac{\partial E}{\partial x} = \frac{q}{\epsilon}(N_d - N_a + p - n) \qquad 2.1.1$$

the charge continuity equations,

$$\frac{\partial p}{\partial t} = \frac{1}{q}\frac{\partial J_p}{\partial x} + \alpha_p|v_p|p + \alpha_n|v_n|n \qquad 2.1.2a$$

$$\frac{\partial n}{\partial t} = \frac{1}{q}\frac{\partial J_n}{\partial x} + \alpha_p|v_p|p + \alpha_n|v_n|n \qquad 2.1.2b$$

and the approximated current density equations,

$$J_p = pqv_p \qquad 2.1.3a$$

$$J_n = -nqv_n \qquad 2.1.3b$$

where E is the electric field, q the absolute value of electronic charge, ϵ the dielectric constant, N_d and N_a the donor and acceptor densities, n and p the electron and hole densities, J_p and J_n the hole and electron current densities, α_p and α_n the ionisation rates of holes and electrons, and v_p and v_n the velocities of holes and electrons.

Since thermal generation provides an insignificant contribution to the total current density, equations 2.1.2. a,b take account of avalanche multiplication only. In addition to this simplification, two further approximations will be made, namely that ionisation rates and scattering limited drift velocities

for holes and electrons in silicon are equal. While this is acknowledged to be erroneous, the simplification aids an analytical approach enabling the physical significance of IMPATT operation to be understood more readily and still yields a realistice dynamic characterisation.

Addition of equations 2.1.2a and 2.1.2b using substitutions from equations 2.1.3 a, b gives

$$\frac{1}{v}\left[\frac{\partial J_p}{\partial t} + \frac{\partial J_n}{\partial t}\right] = \frac{\partial J_p}{\partial x} - \frac{\partial J_n}{\partial x} - 2qv\alpha(n+p) \qquad 2.1.4$$

where $v = |v_n| = |v_p|$, and $\alpha = \alpha_p = \alpha_n$.

Integrating equation 2.1.4 with respect to x from x=o to $x=x_a$

$$\tau_a \frac{dJ}{dt} = \left[J_p - J_n\right]_o^{x_a} + 2J\int_o^{x_a} \alpha\, dx \qquad 2.1.5$$

where $\tau_a = \frac{x_a}{v}$ and $J = J_n + J_p$

From App. B it will be realised that the electron current at x=o consists entirely of reverse saturation current J_{ns} ; similarly at $x = x_a$ the hole current component is simply J_{ps} , the reverse saturation value. Hence the boundary conditions can be expressed as

x=o,

$$J_p - J_n = -2J_n + J = -2J_{ns} + J \qquad 2.1.6a$$

$x=x_a$,

$$J_p - J_n = 2J_p - J = 2J_{ps} - J \qquad 2.1.6b$$

Substituting 2.1.6a, b into 2.1.5 ,

$$\frac{dJ}{dt} = \frac{2J}{\tau_a}\left[\int_0^{x_a}(\alpha\, dx) - 1\right] + \frac{2J_s}{\tau_a} \qquad 2.1.7$$

where the reverse saturation current density $J_s = J_{ps} + J_{ns}$

This equation may be further simplified by neglecting the insignificant term $\frac{2J_s}{\tau_a}$, and substituting a mean value of ionisation rate $\bar{\alpha}$ obtained from evaluating the ionisation integral over the avalanche region. Hence

$$\frac{dJ}{dt} = \frac{2J}{\tau_a}(\bar{\alpha} x_a - 1) \qquad 2.1.8$$

[This is consistent with the steady state condition: when the time derivative of current is zero, $\bar{\alpha} x_a = 1$.]

A small signal perturbation is now applied such that

$$\bar{\alpha} = \bar{\alpha}_o + \bar{\alpha}_a e^{j\omega t} = \bar{\alpha}_o + \bar{\alpha}'_a E_a e^{j\omega t} \qquad 2.1.9$$

$$J = J_o + J_a e^{j\omega t} \qquad 2.1.10$$

$$E = E_o + E_a e^{j\omega t} \qquad 2.1.11$$

where $\bar{\alpha}_a$, J_a and E_a are small signal quantities; $\bar{\alpha}_o, J_o, E_o$ are steady state components; and $\bar{\alpha}'_a = \frac{d\bar{\alpha}_a}{dE}$ enables the substitution $\bar{\alpha}_a = \bar{\alpha}'_a E_a$

Equation 2.1.8 becomes

$$\frac{dJ}{dt} = \frac{2\left[J_o + J_a e^{j\omega t}\right] x_a \bar{\alpha}'_a E_a e^{j\omega t}}{\tau_a} \qquad 2.1.12$$

Neglecting products of small a.c. terms this becomes

$$\frac{dJ}{dt} = \frac{2 J_o x_a \bar{\alpha}'_a E_a e^{j\omega t}}{\tau_a} \qquad 2.1.13$$

or

$$J_a = \frac{2 J_o x_a \bar{\alpha}'_a E_a}{j\omega \tau_a} \qquad 2.1.14$$

This is an expression for the a.c. component of the avalanche conduction current density. A displacement current density J_d will also exist and the total a.c. current density J_t can be defined as

$$J_t = J_a + J_d \qquad 2.1.15$$

where $J_d = j\omega\varepsilon E_a$.

Under given field conditions, the avalanche current J_a is reactive and varies inversely with ω as in an inductor. Similarly J_d is reactive, however it varies directly with ω as in a capacitor. In consequence the avalanche region operates like an L-C parallel circuit where

$$L_a = \frac{\tau_a}{2 J_o \bar{\alpha}'_a A} \qquad 2.1.16$$

$$C_a = \frac{\varepsilon A}{x_a} \qquad 2.1.17$$

where A is the diode cross sectional area.

A resonant frequency f_a for this combination is given by

$$f_a = \frac{\omega_a}{2\pi} = \frac{1}{2\pi\sqrt{LC}}$$

$$= \frac{1}{2\pi}\sqrt{\frac{2 \bar{\alpha}'_a v J_o}{\varepsilon}} \, . \qquad 2.1.18$$

The avalanche zone impedance Z_a may now be deduced since

$$Z_a = \frac{j\omega L_a \cdot \frac{1}{j\omega C_a}}{j\omega L_a + \frac{1}{j\omega C_a}} \qquad 2.1.19$$

which by simple reduction becomes

$$Z_a = \frac{1}{j\omega C_a} \cdot \left[\frac{1}{1 - \frac{\omega_a^2}{\omega^2}} \right] \qquad 2.1.20$$

Within the drift zone, the alternating conduction current $J_c(x)$ can be assumed to propagate as an unattenuated wave at the saturated drift velocity.

Hence

$$J_c(x) = J_a e^{-\frac{j\omega(x - x_a)}{v}} \qquad 2.1.21$$

Since $J_c(x)$ can be related to the total alternating current density by $J_c(x) + J_d(x) = J_t$, equation 2.1.21 may be expressed as

$$J_c(x) = \gamma J_t e^{-\frac{j\omega(x - x_a)}{v}} \qquad 2.1.22$$

where $\gamma = \frac{J_a}{J_t}$.

Using equations 2.1.15, 2.1.21 and 2.1.22, a relationship between the a.c. electric field E_d and current density in the drift region J_t can be derived

$$E_d = J_t \cdot \frac{1 - \gamma e^{\frac{j\omega(x - x_a)}{v}}}{j\omega\varepsilon} \qquad 2.1.23$$

Substitution from 2.1.22 into 2.1.23 and integration over

the drift length $(W - x_a)$ gives

$$V_d = \frac{W - x_a}{j\omega\varepsilon} J_t \left[1 - \frac{1}{1 - \frac{\omega^2}{\omega_a^2}} \left(1 - \frac{e^{-j\theta_d}}{j\theta_d}\right)\right] \qquad 2.1.24$$

where V_d is the a.c. voltage across the drift region, and θ_d is the transit angle of the drift region defined as

$$\theta_d = \frac{\omega(W - x_a)}{v} = \omega \tau_d \qquad 2.1.25$$

with τ_d the drift transit time $= \frac{W - x_a}{v}$

By defining a drift zone capacitance $C_d = \frac{A\varepsilon}{W - x_a}$, it is possible to obtain a complete impedance value Z_d for the drift region:

$$Z_d = \frac{V_d}{J_t A}$$

$$= \frac{1}{\omega C_d}\left[\frac{1}{1 - \frac{\omega^2}{\omega_a^2}}\left(\frac{1 - \cos\theta_d}{\theta_d}\right)\right] + \frac{j}{\omega C_d}\left[-1 + \frac{1}{1 - \frac{\omega^2}{\omega_a^2}}\left(\frac{\sin\theta_d}{\theta_d}\right)\right] \qquad 2.1.26$$

It will be noted that the real component of drift zone impedance is positive for all frequencies below ω_a, and negative thereafter except $Z_d = 0$ when $\theta_d = 2\pi \times$ integer.

A summation of equations 2.1.20 and 2.1.26, together with a passive resistance (R_s) of the inactive region yields an expression for the complete device impedance (Z),

$$Z = \frac{(W - x_a)^2}{2A\varepsilon v}\left[\frac{1}{1 - \frac{\omega^2}{\omega_a^2}}\right]\left[\frac{1 - \cos\theta_d}{\frac{\theta_d^2}{2}}\right] + R_s$$

$$+ \frac{j}{\omega C_d}\left[\left(\frac{\sin\theta_d}{\theta_d} - 1\right) - \left(\frac{\frac{\sin\theta_d}{\theta_d} + \frac{x_a}{W - x_a}}{1 - \frac{\omega_a^2}{\omega^2}}\right)\right] \qquad 2.1.27$$

Gilden and Hines translated this equation into diagrammatic form for small transit angles, and a similar graph appears in Fig. 2.1.2.

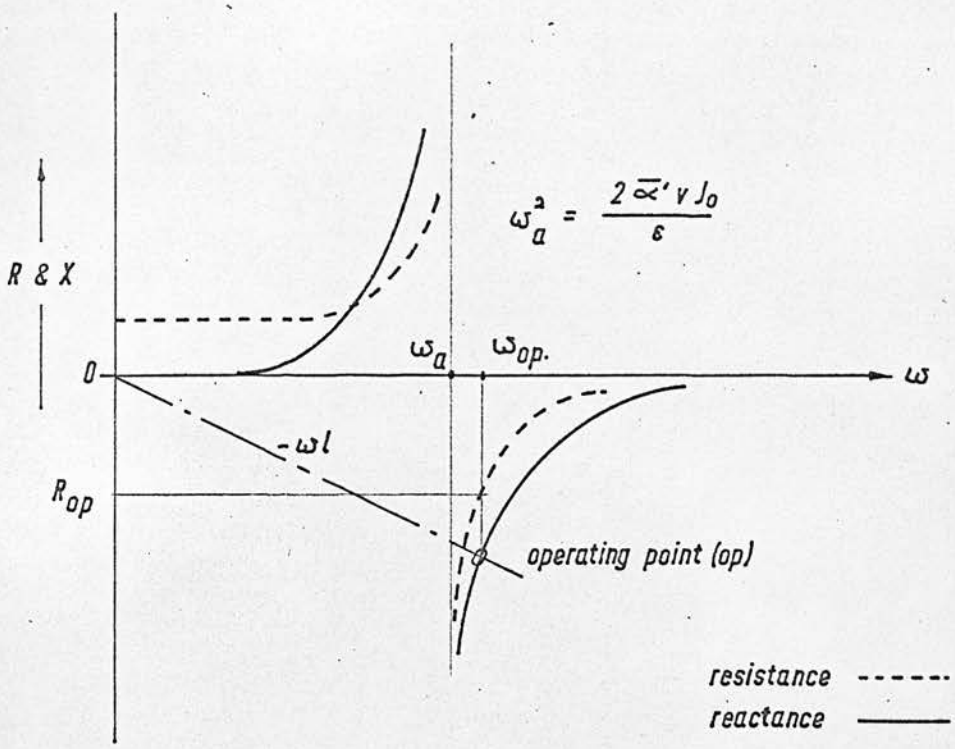
The diode reactance increases with frequency becoming strongly inductive just below the avalanche frequency (ω_a), while its resistance is positive within this range. Above the avalanche frequency, diode reactance becomes capacitive and the resistance becomes negative.

A condition for oscillation may be seen if a negative inductive reactance value is plotted on the same graph. Intersection with the diode reactance curve defines the point where circuit reactance vanishes, i.e. the circuit resonant frequency. A value of diode negative resistance at this frequency can be ascertained by projection to the resistance curve as indicated.

It is worth noting that higher values of series inductance will give rise to larger negative resistance; the operating frequency is always above ω_a, the avalanche frequency; and that an operating frequency can be varied by 'electronic tuning' since avalanche frequency is proportional to the square root of diode current.

(ii) A Large Signal Treatment

While the previous diode analysis describes an IMPATT mechanism and the associated device impedance, free running oscillator applications require an understanding of



Graph of Frequency Dependent Real (R) and Imaginary (X) Parts of IMPATT Diode Impedance

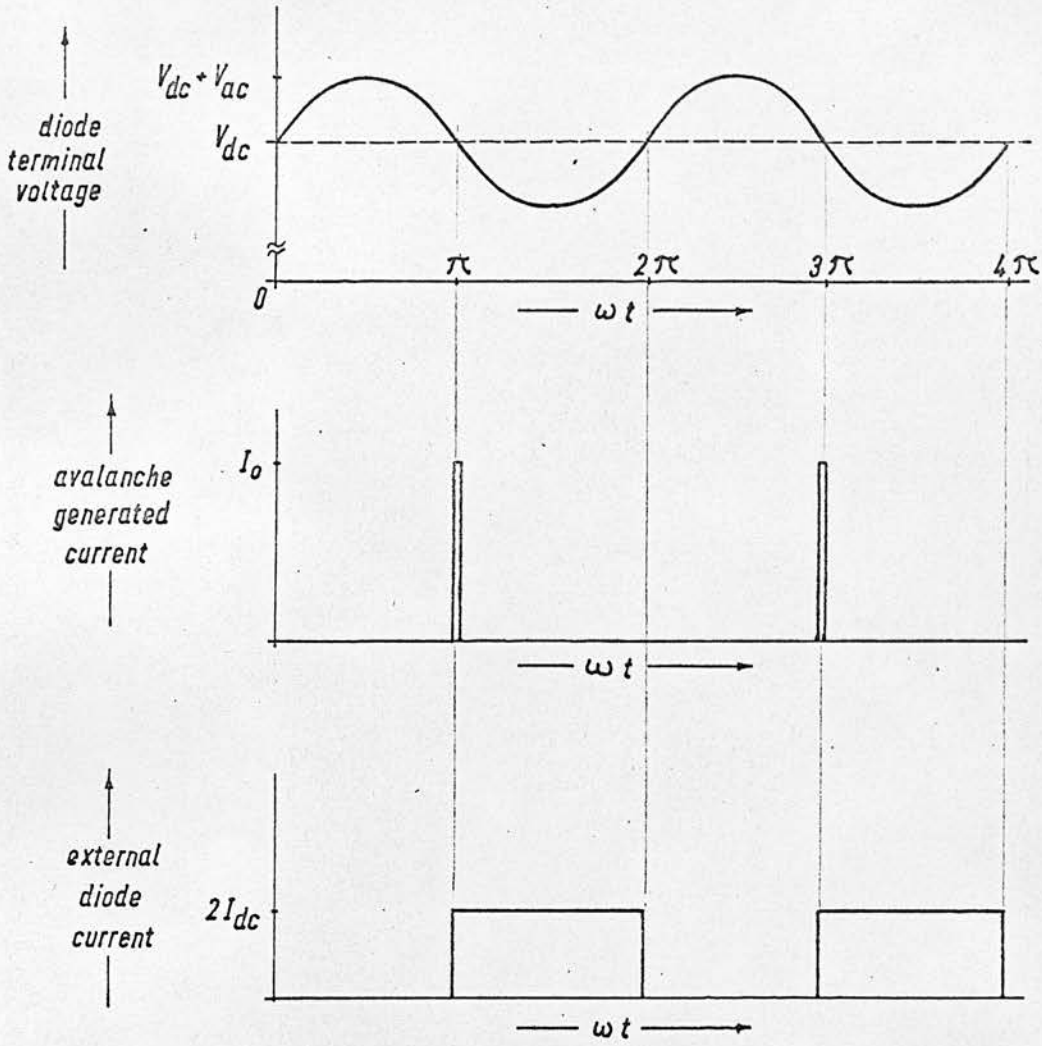
Fig. 2.1.2

large signal performance where sinusoidal variations in current amplitude become unrealistic.

Read considered incremental current levels corresponding to oscillation growth, and provided a theory for the limiting case where carriers are generated in a pulse in the middle of an a.c. voltage cycle. Since small changes in diode voltage and hence electric field can cause large variations in carrier multiplication, it was proposed that a sinusoidal voltage signal could give rise to instantaneous carrier injection as shown in Fig. 2.1.3, a particular case of the general situation depicted earlier in Fig. 1.2.2. The carrier pulse results in an external a.c. particle current which persists during charge transport through the drift region. Since the average particle current is the d.c. bias current (I_{dc}), a maximum signal excursion from zero to $2I_{dc}$ must represent the 100 percent modulation limit.

In accordance with Poisson's Law, the presence of a space charge in the drift region alters the electric field; hence efficient diode operation demands a current density restricted by the magnitude of charge pulse which can travel through the drift space without reducing the electric field to a level below that required for saturated scattering limited carrier transit velocity v_{sl}.

Total bias current is limited by the maximum permissible device cross sectional area (πR^2), since the resultant reverse biased diode capacitance (C) must be matched by a circuit



Large Signal Generation Sequence

Fig. 2.1.3

inductance (L) such that the impedances equate at the resonant frequency $(\omega_{op} = \frac{\pi v_{sl}}{W})$; where W is the junction depletion width. For silicon, this capacitance can be expressed as

$$C = \frac{3.33 R^2}{W} \times 10^{-12} \qquad 2.1.28$$

with units of C in farads, W in cm., R in cm.

Hence the impedance condition :-

$$\frac{10^{12} W^2}{3.33 \pi v_{sl} R^2} = \omega_{op} L \qquad \text{where } L \text{ is in henries}$$

i.e. $$\left(\frac{R}{W}\right)^2 = \frac{1.37 \times 10^4}{\omega_{op} L} \qquad \text{for } v_{sl} = 7 \times 10^6 \text{ cm.s}^{-1}$$

For a practical lower limit of $\omega_{op} L = 10$ ohms, the diode radius must not be greater than 37 times the depletion width.

A diode efficiency (η) may be defined as the ratio of the average a.c. power delivered $(\overline{P_{ac}})$ to the d.c. power supplied (P_{dc}) where

$$P_{dc} = V_{dc} I_{dc}$$

$$\overline{P_{ac}} = \frac{1}{2\pi} \int_0^{2\pi} V_{ac} \sin \omega t . \widetilde{I_{ac}} \, d(\omega t)$$

$$\widetilde{I_{ac}} = 0 \Big|_{\omega t = 0}^{\omega t = \pi} , \quad \widetilde{I_{ac}} = 2 I_{dc} \Big|_{\omega t = \pi}^{\omega t = 2\pi}$$

Therefore :-

$$\eta = \frac{2|V_{ac}|}{\pi V_{dc}} \qquad 2.1.29$$

$$= \frac{2|V_{ac}|}{\pi(V_a + V_d)} \qquad 2.1.30$$

where V_a and V_d are the avalanche and drift region d.c. voltages.

Read proposed a value $|V_{ac}| = \frac{V_d}{2}$ which would not seriously affect the transit velocity of carriers, to achieve a theoretical device efficiency of $\frac{1}{\pi}$ or approximately 30 percent in the limiting case where $V_a \ll V_d$.

Recognising a dissimilarity in ionisation rates for electrons and holes in silicon such that the electron ionisation integral with distance is not unity but about three, Gummel and Scharfetter considered further Read's postulate of insignificant voltage drop across the avalanche region to propose a reduced theoretical maximum diode efficiency.

If $\bar{E_a}$ and $\bar{E_d}$ are the average electric fields in the avalanche and drift regions, while x_a and x_d are the effective avalanche and drift widths,

$$V_a = \bar{E_a} x_a \qquad 2.1.31$$

$$V_d = \bar{E_d} x_d \qquad 2.1.32$$

By defining the drift frequency $f_d = \frac{v_{sl}}{2x_d}$, the drift voltage can be expressed as

$$V_d = \frac{\bar{E}_d v_{sl}}{2 f_d} \qquad 2.1.33$$

If the charge initiating a.c. voltage amplitude $|V_{ac}| = \frac{V_d}{2}$ then from Gauss' Theorem, the generated pulse $Q\ (= \varepsilon E) = \frac{\varepsilon \bar{E}_d}{2}$.

Under 100 percent current modulation, the d.c. bias current density (J_0) will be

$$J_0 = \frac{\varepsilon \bar{E}_d f_d}{2} \qquad 2.1.34$$

Using the approximations that ionisation rate varies as the sixth power of electric field[12], and that the integral of electron ionisation rate is about 3, i.e.

$$\bar{\alpha}_a x_a = 3 \qquad 2.1.35$$

$$\bar{\alpha}'_a \simeq \frac{6 \bar{\alpha}_a}{\bar{E}_a} \qquad 2.1.36$$

it is possible, using equations 2.1.31 - 2.1.36, to rewrite equation 2.1.18 in the form

$$(2\pi f_a)^2 = 2 \times 6 \times 3 \times \frac{V_d f_d^2}{V_a} \qquad 2.1.37$$

or

$$\frac{V_d}{V_a} = \left(\frac{f_d}{f_a}\right)^2 \qquad 2.1.38$$

Despite simplifying assumptions, Gummel and Scharfetter were able to conclude that a diode designed for efficient operation at frequencies close to the avalanche frequency ($f_d = f_a$) will have similar voltages across the avalanche and drift regions. As a result, Read's efficiency prediction was modified from equation 2.1.30 to give a theoretical maximum of 15 percent in silicon.

The validity of this simple analysis was verified by complex numerical computation where consistent solutions for carrier generation, carrier transport, and charge continuity equations were calculated to describe the progressive electrical behaviour of the IMPATT diode and its associated resonant circuit. In determining the evolution of electron and hole concentration, electric field, and terminal voltage and current at various instants during the a.c. steady state oscillation, Gummel and Scharfetter took account of carrier recombination, diffusion currents, mobility dependence on electric field and ionised impurity density, and dissimilarities in electron and hole properties. This rigorous numerical treatment yielded diode efficiencies of 9 to 18 percent from an order of magnitude variation in d.c. current density, and demonstrated further that the square wave approximation for large signal currents must be modified slightly to include the extra component of particle current which flows while the holes generated with the electrons are swept into the p^+ region.

Further numerical analyses of large signal IMPATT diode behaviour have been performed under the direction of Professor E. A. Ash at Imperial College London, however publication of this work is restricted to C.V.D. reports which, according to the research contract, may be obtained by application to the issuing authority.

In an ideal thermal environment the diode, placed in intimate contact with the end wall and central conductor of a coaxial cavity, can be considered sufficiently thin and small in diameter that temperature gradients within the silicon are negligible and heat will flow into the surrounding metal in all directions. The temperature difference (ΔT) between the diode and outer cavity surface may be calculated from a standard heat transfer equation(54)

$$\Delta T = \frac{\pi R P_h}{8K} \qquad 2.1.37$$

where R = diode radius.

K = thermal conductivity of the surrounding metal.

P_h = power density responsible for heat generation.

For a given power density therefore, diode temperature will increase linearly with radius, and at lower frequencies the impedance restriction shown earlier to confine $R \leq 37W$ will be secondary to thermal limitations.

While quantitative theoretical interpretation of this physical situation is extremely complex (since carrier generation and recombination, diffusion currents and carrier velocities are all temperature dependent) practical comparisons on the use of different heat sinking materials have been undertaken by Decker et al.[55] and Swann[48]: further reference to this work will be made during the experimental development of an appropriate laboratory mounting configuration.

2.2 DESIGN OF AN ABRUPT JUNCTION P^+NN^+ IMPATT DIODE

It was indicated in Section 1.2 that satisfactory IMPATT operation requires the length of the drift zone to be such that carriers will move across it in one half of the proposed period of oscillation.

From Section 1.3 and Fig. 1.3.1 b, the approximation can be made that drift and depletion lengths are equal since ionisation is extremely localised, and that carriers travel at constant space charge limited velocity.

In attempting to manufacture a diode capable of oscillation in the 9-10 GHz range, design was centred on a frequency of 9.5 GHz and the appropriate depletion length *(W)* calculated as

$$W = \frac{v_{sl}\, T}{2} = 3.7 \text{ microns.}$$

where

$T = \frac{10^{-9}}{9.5}$ s., the oscillation period.

$v_{sl} = 7.10^{6}\ cm.s^{-1}$, an approximation to the scattering limited velocity for electrons and holes in silicon.

The other important parameter in wafer processing is the n-region doping level, uniquely determined from the drift length and an additional desirable feature that diode breakdown should coincide with complete depletion of the n-region. In this situation the diode series resistance is minimised, and the problem of undepleted high resistivity material discussed by Kovel and Gibbons[11] is eliminated.

Application of classical semiconductor theory to the reverse

biased, one-sided, abrupt junction yields the expression for depletion length W , in terms of impurity concentration N_d and voltage V is a summation of built in and applied voltages:

$$W = \sqrt{\frac{2\varepsilon V}{qN_d}} \qquad 2.2.1$$

Derivation of this relationship is detailed in Appendix A.

It remains to establish the applied voltage at breakdown as a function of impurity level before the n-region can be completely specified. This involves solution of the basic ionisation integrals under avalanche conditions as developed in Appendix B:

$$\int_0^W \alpha_p \, e^{-\int (\alpha_p - \alpha_n) dx} = 1 \qquad 2.2.2a$$

$$\int \alpha_n \, e^{-\int (\alpha_n - \alpha_p) dx} = 1 \qquad 2.2.2b$$

Unfortunately these equations cannot readily be integrated analytically and numerical computation is necessary. Sze and Gibbons [12] used a substitution for the field dependent electron and hole ionisation rates of the form

$$\alpha, \beta = A e^{-\left[\frac{b}{E(x)}\right]^m} \qquad 2.2.3$$

where A, b and m are constants to convert equations 2.2.2a, b to functions containing field and depletion width variables. Further substitution of Poisson's equation with appropriate electric field boundary conditions provided an expression

for depletion width at breakdown which could be solved at any value of impurity concentration by iterative method. Cumulation of such results enabled Sze [13] to generate a graph of depletion width at breakdown as a function of impurity concentration and this has been reproduced in Fig. 2.2.

Furthermore, the maximum electric field strength (E max) can be expressed as

$$-E_{max} = \left.\frac{dV}{dx}\right|_{x=0} = \frac{qN_dW}{\varepsilon} \qquad 2.2.4$$

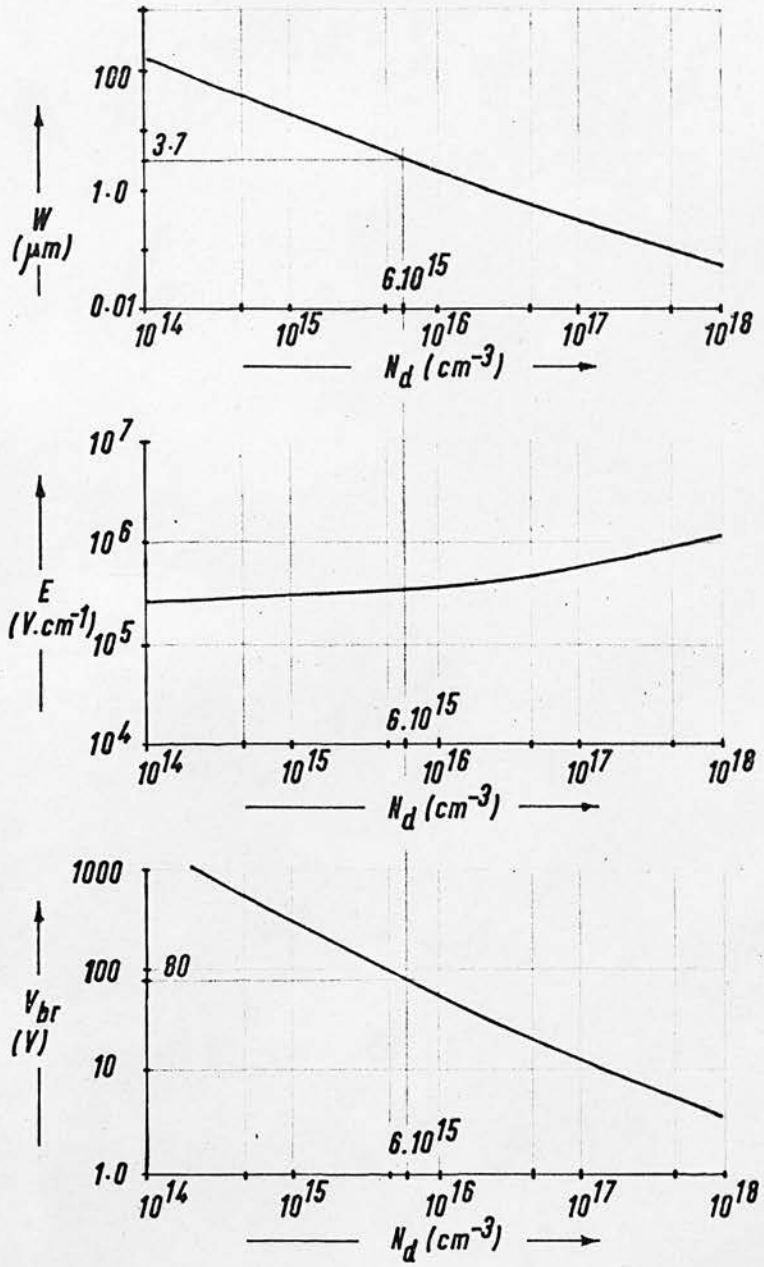
and hence the breakdown voltage (V_{br}):

$$V_{br} = \frac{\varepsilon E_{max}^2}{2qN_d} \qquad 2.2.5$$

These equations are also represented graphically in Fig. 2.2 from which it can be estimated that a depletion length of 3.7 microns will occur in the n-region of an abrupt silicon p^+n junction at a reverse breakdown voltage of 80 volts providing the n-type impurity concentration is 6×10^{15} atoms.cm^{-3}.

In addition to calculating the silicon wafer requirements, it is necessary to determine a useful range of junction areas for the device. Circuit impedance restrictions dictate a maximum limit, and areas from 0.2×10^{-4} cm^2 to 5×10^{-4} cm^2 will provide reactances within the range 300 ohms to 15 ohms.

Apart from these almost statutory design requirements,



Graphs of Depletion Width at Breakdown (W), Maximum Electric Field Strength (E), and Breakdown Voltage (V_{br}) against Impurity Concentration (N_d).

Fig. 2.2

other parameters must be considered e.g. thermal dissipation. Since they are largely dependent on process technology, discussion will be deferred until later in the more practical context of device fabrication.

CHAPTER 3

SELECTION OF A DIODE FABRICATION PROCESS

The problem of processing silicon to approach the idealised doping profile described in the preceding chapter was considered in the light of currently documented technologies; and in addition, thought was given to alternative techniques which might yield a more attractive diode in terms of superior performance or less costly fabrication.

Methods of tailoring impurity distribution in a silicon slice may be subdivided as follows:

(1) Modification of the basic slice to alter an original doping level. This may be carried out by DIFFUSION or IMPLANTATION.

(2) Extension of the basic slice by means of single crystal silicon growth on the parent surface, during which impurity dopant is added to provide the requisite profile. This is the technique of EPITAXY.

(3) Surface combination of a parent slice with impurity dopant during the regrowth cycle of an ALLOYING operation.

Each of these techniques has fundamentally different equipment and process requirements, each has its own limitations

in application, and each gives rise to a characteristic doping profile. It is the purpose of this chapter to indicate where single and complementary processes may be used to fabricate the required three layer structure, and to demonstrate the relative merits of proposed experimental diode configurations.

Although alloying was the first used method of semiconductor doping during early work on solid state diodes and transistors, the process was largely neglected as effort directed towards a solid state diffusion technique which has subsequently become a fundamental operation in modern monolithic integrated circuit technology. A similarly high degree of control has been obtained from the epitaxial growth process which progressed in sophistication to complement diffusion in the fabrication of semiconductor devices.

Since these two techniques are accepted manufacturing operations within the microelectronics industry, it seemed reasonable to consider their application to the problem of IMPATT diode fabrication, before examining alternative methods.

3.1 DIFFUSION INTO EPITAXIAL SILICON

3.1.1 EPITAXIAL GROWTH OF N ON N^+ SILICON

In the past, the most popular epitaxial deposition procedure has been to chemically reduce silicon tetrachloride ($Si\ Cl_4$) in a hydrogen carrier gas within a reactor where the silicon slice

is heated to over 1150°C. Under ideal conditions this will result in an extension of the parent crystal as silicon is deposited from the $SiCl_4$ vapour phase. Variation in deposition rate is achieved by altering the mole fraction of $SiCl_4$ in the gas stream, by control of total flow rate, and by adjustment of the silicon slice temperature. The relationships between these parameters are described by Theurer[14] in his article on $SiCl_4$ epitaxy, and further details of the technology are provided in a review paper by Cave and Czorny[15].

Although a wide range of impurity species can be added to the reaction to yield p and n layers over a wide range of resistivity, deposition of high resistivity material is limited by diffusion of impurities from the heavily doped substrate into the growing layer to modify the ultimate profile. This 'out-diffusion' mechanism has been discussed by Grove et al[16] and Kahng[17], while Rice[18] provided analytical data from which an accurate prediction of impurity distribution can be made.

A further undesirable effect which must be accomodated is that of 'auto-doping', which involves the transport of impurities from the reverse surface of the silicon slice to the growing face where deposition may take place. This problem, considered by Grossman[19] and later by Gupta and Yee[20] can be eliminated by sealing the underside of the sample with a pyrolytically grown oxide.

Although movement of impurity atoms from the parent slice

into the high resistivity layer will constitute a deviation from the ideal abrupt structure, such an effect may be minimised by suitable choice of low mobility substrate dopant, and by optimising time and temperature parameters of the epitaxial deposition. When a combination of epitaxy and diffusion is employed however, the merit of developing and abrupt n n^+ junction in the initial fabrication stage is questionable, since subsequent diffusion involves a protracted high temperature cycle when attendant out-diffusion from the substrate is inevitable.

3.1.2 DIFFUSION OF P^+ IMPURITY INTO AN EPITAXIAL LAYER

The diffusion process is commonly performed in an open furnace tube capable of temperatures in excess of 1250°C, controlled by suitable profiling to $\pm \frac{1}{2}$°C over the reaction zone length. Conventional diffused junction formation is a two stage process, involving deposition of an impurity source on the silicon slice and subsequent 'drive in' where diffusion is performed to the correct depth according to device requirements.

In order to minimise out-diffusion of the n^+ substrate during a high temperature cycle, it is desirable to use a dopant exhibiting high diffusivity which will enable the correct profile in the shortest possible time. It is fortuitous that boron, the only practical p-type impurity, has such a property.

Initially, the slice is heated to about 1150°C in an inert gas stream such as nitrogen. During impurity deposition, an

additional nitrogen stream is passed over a boron tribromide (BBr_3) liquid dopant source, and BBr_3 vapour carried into the furnace tube is decomposed by the inclusion of oxygen in the system. A 'local source' of oxide (B_2O_3) is thereby obtained within the reaction zone where the silicon surface oxidises yielding elemental boron and a glassy borate layer. For the immediate future, this layer can be considered as an infinite source of impurity from which diffusion into the silicon crystal commences. Studies by Williams [21] and Thurston[22] have indicated that over the range of background impurity concentration 10^{14} -10^{17} atom. cm.$^{-3}$, this diffusion obeys a complementary error function of the form

$$C(x) = C_s \operatorname{erfc}\left[\frac{x}{2\sqrt{Dt}}\right] \qquad 3.1.2.1$$

where $C(x)$ is the impurity concentration at a distance x from the surface after a time t.

C_s is the surface impurity concentration.

D is the diffusion coefficient.

Normal diffusion practice is to arrest this process after a few minutes, and remove the boron glass prior to a higher temperature diffusion stage utilising the very shallow, high concentration impurity in the vicinity of the silicon surface. While this additional step is beneficial in integrated circuit and transistor manufacture, enabling corrections to be made for earlier inaccuracies and providing a generally more applicable Gaussian dopant profile from the limited impurity source, it was

rejected for present work in favour of single stage deposition and diffusion giving rise to more abrupt junctions.

3.1.3 AN EXPERIMENTAL DIODE DESIGN

Having discussed the two distinct stages in silicon processing, it is necessary to determine a practical structure, in order that later comparison may be made with other techniques.

3.1.3.1 SELECTION OF A SILICON SUBSTRATE

The characteristics of silicon used as a starting material have significant consequences during later stages of device fabrication. Bean and Gleim[23] considered many aspects of silicon technology, and the processing constraints imposed by crystal orientation.

Although single crystal silicon has a hexa-octahedral structure and may be grown with (111), (110) or (100) orientations, material of the (111) type is most commonly used for reasons of easy growth and reproducibility. Since the (111) plane has the highest atomic packing density, silicon grown in the ⟨111⟩ direction is least susceptible to edge fracture during any mechanical polishing. Nevertheless, the (111) plane exhibits the slowest growth rate and is similarly slow to etch. Epitaxial deposition therefore requires that the substrate be cut at least 1.5° off the (111) plane towards the nearest (110).

By comparison, (100) orientated material is basically

suited to epitaxy and has been shown by Bean and Gleim to allow faster boron diffusion than (111) type. This will permit a reduced time-temperature cycle during junction formation, thereby minimising substrate out-diffusion. An additional advantage of (100) material is the definition obtainable when chemically etching patterns using surface masks. The ability to etch preferentially in a vertical direction proves extremely useful during mesa formation when device profiling can be controlled throughout a prolonged etching cycle.

Similar etching properties are observed in (110) orientated silicon, however the impurity diffusion rate is lower and comparable with (111) material. Being inherently more difficult to grow and having no unique advantages in our application, (110) silicon will receive no further attention.

From electrical considerations, the substrate will ultimately form the n^+ region of the device, and for high operating efficiency must constitute a low resistance contact. An essential requirement therefore, is a high degree of impurity doping, and thought must be given to the choice of a suitable element in view of future slice processing. Table 3.1.3.1 extracted from Wolf[24] lists diffusion coefficients for dopants currently used in silicon wafer manufacture.

	1300°C	1200°C	1000°C
Phosphorus	10^{-11}	3×10^{-12}	0.5×10^{-13}
Arsenic	2×10^{-12}	2×10^{-13}	10^{-15}
Antimony	2×10^{-12}	3×10^{-13}	2×10^{-15}

Table 3.1.3.1 Impurity Diffusion Coefficients ($cm^2\ s^{-1}$)

It is considered that antimony (Sb), or arsenic (As) should be selected to minimise out-diffusion during epitaxial deposition and subsequent diffusion. A suitable specification for the silicon substrate would include (100) orientation and Sb doping level > 10^{19} atom. cm^{-3}, in addition to dimensional requirements to be considered later.

3.1.3.2 DESIGN OF THE EPITAXIAL LAYER

The epitaxial region defines critical operating parameters of the ultimate device, and relationships between junction depletion width, breakdown voltage and impurity doping level have been defined in Section 2.2. While it was established from dynamic considerations that a depletion length of 3.7 μm is necessary, the epitaxial layer must also support p^+ diffusion which will complete the diode structure. In order to suppress substrate out-diffusion, an effort must be made to minimise deposition time and the subsequent diffusion schedule. A total epitaxy thickness of 5.7 μm will allow rapid p^+n junction formation at a depth of 2 μm: this will be confirmed during discussion of an experimental profile.

The remaining consideration is the choice of dopant. Since arsenic is readily available in a gaseous arsine (AsH_4) form for epitaxial work, it is though suitable, and the diffusivity although low is an insignificant property since layer boundaries are controlled by the adjacent heavily doped regions.

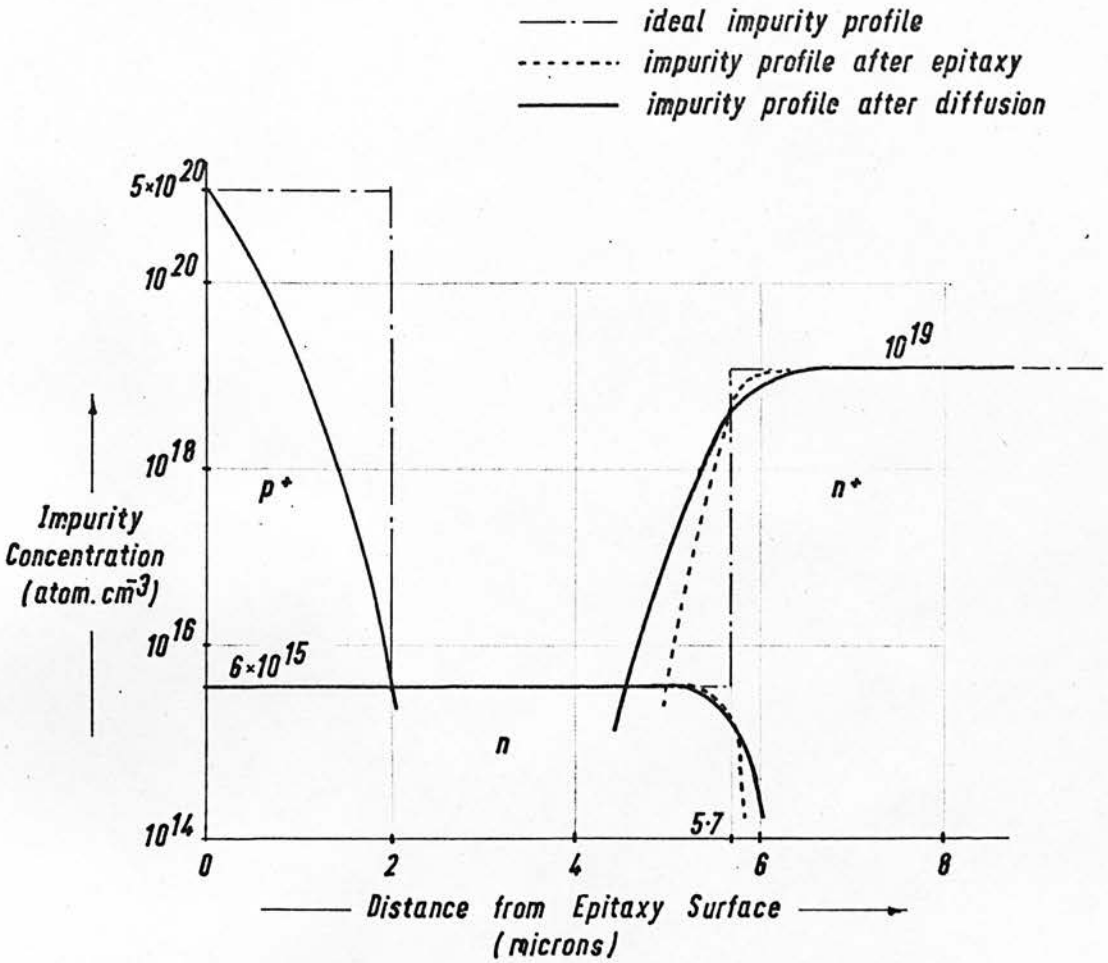
3.1.3.3 DESIGN OF THE P^+ DIFFUSION CYCLE

The objective of p^+ diffusion is to achieve an abrupt one-sided junction with the epitaxially grown n-layer. As discussed in Section 3.1.2, the optimum technique involves retention of an 'infinite' p-type impurity source invariant at about 5×10^{20} atom. cm^{-3}, during diffusion to a junction depth determined by the time temperature cycle. In this case, the electrical junction must occur 2.0 μm from the epitaxial surface.

Although the original source material may be varied according to system requirements, the BBr_3 technique described earlier is convenient for this work.

3.1.3.4 THE EXPERIMENTAL PROFILE

Fig. 3.1.3.4 demonstrates the impurity distribution arising from epitaxy and subsequent diffusion performed upon an n-type antimony doped silicon slice. It has been deduced in Appendix C by application of the Rice[18] theory for out-diffusion, and the assumption of complementary error function profile associated with p^+ diffusion.



Experimental Impurity Profile for Diffusion into Epitaxy

Fig. 3.1.3.4

Production of a graded p^+n junction has the effect on device operation of altering the electric field pattern, thereby distributing the multiplication zone. Furthermore, out-diffusion of impurities from the substrate gives rise to a reduction in depletion width at breakdown which may occur before punch through at the original substrate surface. This will result in increasing the residual series resistance of the diode.

If an accurate determination of device dynamic characteristics is required, a more rigorous design study will be necessary to accommodate such deviations from an ideal abrupt profile.

3.2 DOUBLE EPITAXY

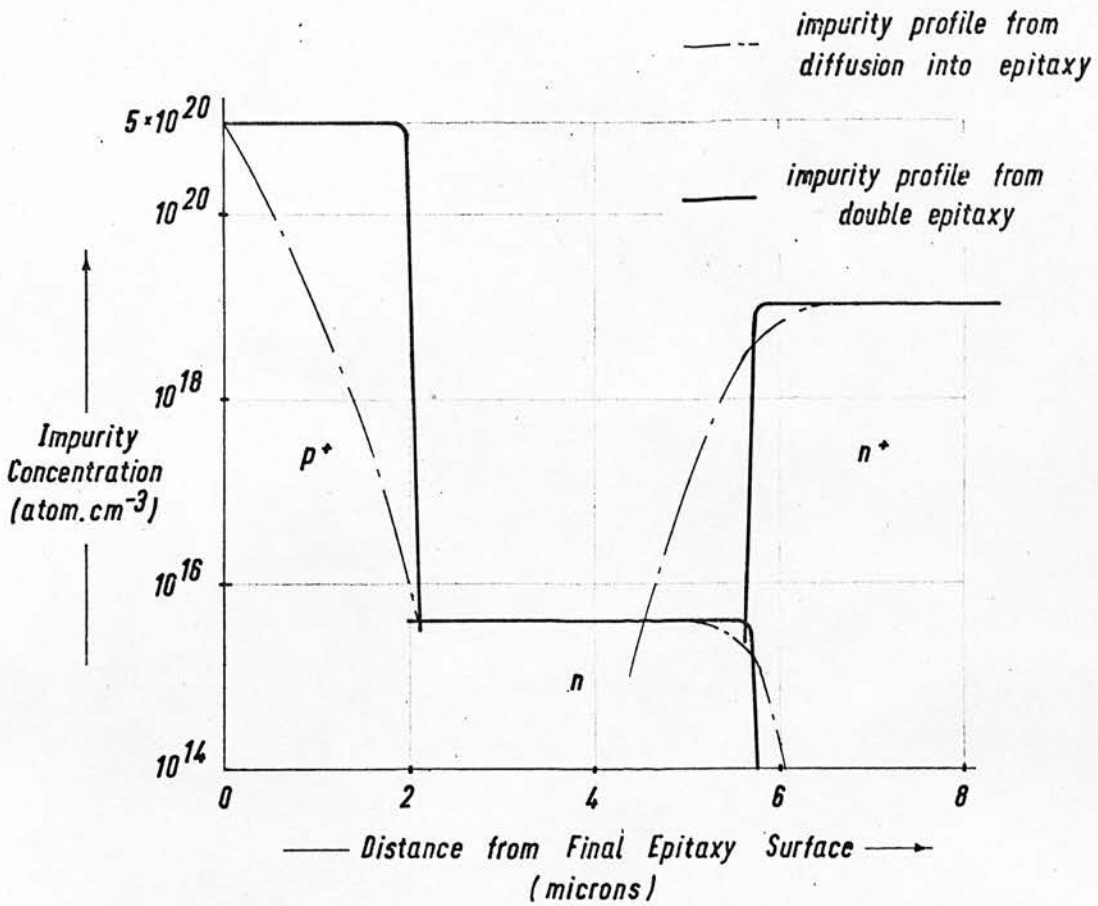
The possibility of fabricating the n and p^+ regions of an IMPATT diode by sequential epitaxial deposition using the hydrogen reduction of silicon tetrachloride has received little attention in the past because of problems associated with the control of different dopant species within the same reactor tube at high temperature. Furthermore, successful double epitaxy of this nature could only result in an impurity profile comparable with that already available from diffusion into single epitaxy.

In 1963, Joyce and Bradley[25] showed that direct pyrolysis of silane (SiH_4) enabled silicon epitaxy at temperatures as low as 1000°C. The principle which they established in a complex, low pressure experiment was subsequently exploited by Bhola and Mayer[26], using equipment similar to that already developed

for $SiCl_4$ reduction, to demonstrate an improvement in abruptness of profile resulting from reduced substrate out-diffusion and the elimination of autodoping from a process which would not generate any reactive products during deposition. Further work by Gupta and Yee[20] confirmed these results, although autodoping which they observed had to be suppressed by the substrate sealing procedure mentioned earlier.

Attempts to reduce epitaxy system temperatures further proved unsuccessful because of a growth inhibiting mechanism attributed to the desorption of hydrogen during SiH_4 dissociation at the silicon surface. Hydrogen evolution provided a localised gaseous medium through which incoming SiH_4 had to diffuse before useful pyrolysis could be achieved. Recognising this restriction, Richman and Arlett[27] attempted to minimise the desorption rate using a helium carrier gas to reduce the hydrogen partial pressure. In consequence, they obtained epitaxial silicon growth at 800°C, maintaining an acceptable deposition rate of $0.5\,\mu m\,min^{-1}$.

This achievement has stimulated new interest in double epitaxy and $p^+n\,n^+$ structures have been fabricated by Richman et al[28] demonstrating extremely abrupt junction profiles. Although application of this particular construction has not been widely publicised, reference was made by Stover[29] in 1968 to the production of IMPATT diodes by double epitaxy, and an estimated impurity distribution is compared in Fig. 3.2 with earlier results



Comparison of Theoretical Impurity Profiles for Double Epitaxy and Diffusion into Epitaxy

Fig. 3.2

obtained from diffusion into single epitaxy.

3.3 ION IMPLANTATION INTO EPITAXIAL SILICON

3.3.1 INTRODUCTION

The method of introducing impurity atoms into a semiconductor by accelerating them to high velocities and directing them against the crystal surface was originally proposed in a patent filed by Schockley[30] in 1954, which described very shallow impurity doping suitable for high frequency junction transistor fabrication. Despite considerable practical effort by Cussins[31] on implantation in germanium, the first successful device relying on the chemical properties of an implanted ion was not realised until 1961, when Alvanger and Hansen[32] produced a particle detector processed in this way. In recent years, ion implantation has been used to provide diodes, field effect and bipolar transistors, as efforts continue to exploit the advantages of this technology.

While production of a functional device remained the primary objective of our research, the opportunity of contributing to a new discipline could not be ignored. A microelectronics industry heavily committed to diffusion and epitaxy cannot sustain the capital or development costs of such a study, and can only benefit from research directed towards an improved technique for controlled junction formation in silicon. For this reason, considerable time was devoted to analysis of possible ion

implantation systems in an attempt to design the most economical plant to meet the immediate requirements of this project.

3.3.2 BASIC IMPLANTATION TECHNOLOGY APPLIED TO SEMICONDUCTOR PROCESSING

Reduced to its simplest form, an implantation system will consist of an ion source, accelerator assembly and target chamber. The need for further sophistication will become obvious as process limitations are outlined.

The choice of ion source depends largely on the elemental species employed, and the compound in which it is originally present. Wilson [33] has written a comprehensive paper on the relative merits of sources using vapour electron bombardment, sputtered electron bombardment, and surface ionisation, from which it is deduced that the most versatile for semiconductor doping is the vapour electron bombardment type. It provides a high beam current, and may be designed to deliver several different ion species during one processing cycle.

In particular, the r.f. ion source designed by Thoneman[34] and developed by Moak et al.[35] is suitable for use with the corrosive compounds from which Group *III - V* dopants are readily obtainable. A plasma discharge can be initiated with an r.f. coil mounted externally to the source, and useful operating lifetime will then be restricted only by contamination from materials condensed on the interior walls of the ionisation chamber. Ion extraction is achieved by providing metal electrodes

at either end of the discharge tube, both carefully shielded from the plasma and supplied with a voltage differential to initiate ion drift.

Acceleration of ions to the required energy level for implantation is achieved by interaction with an electrostatic field created when two or more apertured conducting plates are placed in a high vacuum environment adjacent to the ion source and supplied with the appropriate voltages.

When an ion enters a single crystal silicon sample at high velocity, it loses energy through collisions with silicon atomic nucleii and electrons, until finally coming to rest. The total distance travelled, called the range, has been the subject of a paper by Gibbons[36], in which he describes the random variables of collision incidence and energy transfer per collision, and the effect that they have an impurity penetration. Statistically, for a given energy, there will exist a mean implant range and a Gaussian distribution about this, which can be characterised by a standard deviation.

This situation is further complicated by a phenomenon known as channelling which affects ions entering the target crystal in a low index crystallographic direction. Some of these ions, travelling midway between atomic planes, will lose little energy in nuclear collisions, and electronic stopping will determine their ultimate range. In other cases, where ions have a shallow incidence angle to the channel direction, they are guided between

adjacent atomic planes by occasional corrective nuclear collisions.

Although the channelling property allows greater sample penetration for a given ion energy, effective exploitation requires precise crystal orientation, a damage free silicon surface, and depends on minimal bulk damage from non-channelled ions. In general, the dosage for damage free implants is restricted to a level insufficient for semiconductor doping, and normal practice is therefore to eliminate the channelling mechanism by suitable alignment of the crystal within the target chamber. Since the direction of ion entry must be well defined, it is necessary to ensure minimum beam dispersion through collisions with residual gas atoms: accordingly the target chamber and any preceding free flight region must have a high vacuum capability.

3.3.3 SYSTEM REFINEMENTS

For our requirements, restricted to the fabrication of a p^+n junction in an existing nn^+ silicon target, it was necessary to obtain a high and uniform doping level over the entire slice. One approach to this problem involves diffusing the ion beam after acceleration, by the introduction of a suitable lens assembly and free flight tube, in order that uniform target irradiation can be achieved. This technique was used by Ferber[37] who indicated that a significant proportion of the ion beam must be discarded with a resultant increase in implantation time for a required impurity level. An alternative and more attractive method of achieving uniformity, described by Martin et al.[38],

involves scanning the beam over the target in a regular manner, thereby utilising most of the incident source material and avoiding the need for beam profiling. For this purpose, orthogonal pairs of electrostatic deflection plates are positioned following the acceleration tube and voltage is supplied from appropriate ramp generators.

Consideration must be given to the elimination of unwanted ion species in the output beam arising from trace impurities, ionised parent molecules of the compounds used as source charge material, and ions of all possible dissociated forms. In addition, neutrals generated by charge transfer between the ion beam and residual gas in the system must be removed. Normal practice is to use a magnetic analysis, where the beam is constrained to pass through a uniform magnetic field causing ions to assume circular paths in a plane at right angles to the field direction. The radius of curvature for a given ion species is determined by its momentum and charge state, together with the magnetic field strength. Trajectories of ions with identical energies are determined uniquely by the charge to mass ratio, and aperture selection will separate the desired species from the rest. Since a neutral beam will remain undeflected in the magnetic field, it will be automatically excluded from the analyser output.

While this separation technique requires a large magnet and power supply in order to achieve satisfactory resolution, economy in size and cost can be obtained by combining electric and magnetic fields in a cross field velocity analyser. Orthogonal

electric and magnetic fields are arranged to provide equal and opposite forces on ions of the desired charge to mass ratio while all other charged species are deflected from axial transit and arrested by a suitably apertured plate. Elimination of the unwanted neutral beam can be achieved by applying electrostatic deflection in advance of the scanning region to which neutrals will not respond.

One of the problems of implantation is the displacement of atoms from the regular silicon crystal lattice by penetrating impurities. The microscopic damage can be annealed out at temperatures between 300°C and 600°C, during which time the impurities become substitutional to a degree dependent on the particular species: in the cases of boron and phosphorus, nearly complete utilisation of the implant is possible. This process can be performed as a post-implantation stage or alternatively, provision for target heating within the system will allow doping and simultaneous damage recovery.

Although our system requirements did not include any method of device delineation, it is appropriate in an appraisal of system refinements to mention the most attractive possibility of programming the ion beam to irradiate defined target areas in the formation of a complex integrated circuit. This feature presents the opportunity of fabricating not only p and n regions within a silicon slice, but also allows the ionic deposition of a conducting interconnection pattern to complete the integrated

circuit in a single and automated operation. Such a technique would require a combination of high beam resolution and high current density which to the present day remains unattainable. A contribution towards solving these problems would vastly improve the status of a potentially superior semiconductor processing technology which has to rely on a series of 'in contact' metallic masks defined using conventional photolithographic and chemical etching methods.

3.3.4 AN EXPERIMENTAL IMPLANTATION SYSTEM

In the design phase, it was necessary to compromise between a rudimentary system which would simply permit p^+ doping of a silicon slice, and a more versatile apparatus capable of development for later studies on multiple and programmed implants. Fig. 3.3.4.1 shows the system schematic for an equipment suitable for present purposes and amenable to later sophistication. It incorporates an ion source of the r.f. type described by Moak, which is compatible with boron, phosphorus, and arsenic deriving compounds. Continuous operation for 50 hours is not unreasonable, and a beam current in excess of 10 μA. can be maintained throughout. A 30° magnetic analyser has been included to utilise equipment currently available in our laboratory. While this is adequate for low energy boron implantation, greater penetration or use of higher mass ions would necessitate a replacement. Power supply configuration has been chosen to allow the target chamber to be at earth potential, thereby facilitating sample manipulation,

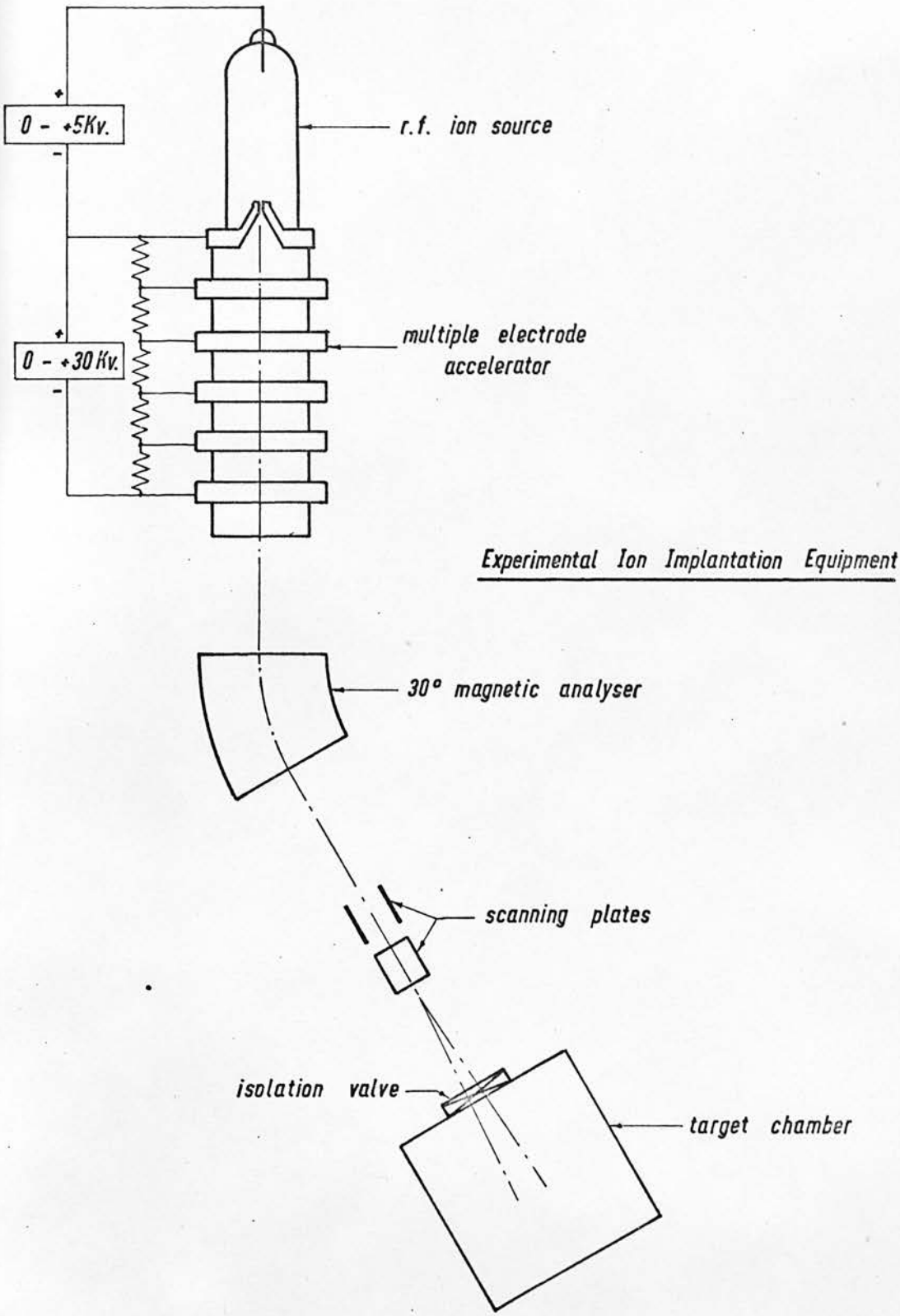

0 - +5Kv.
r.f. ion source
0 - +30Kv.
multiple electrode accelerator
Experimental Ion Implantation Equipment
30° magnetic analyser
scanning plates
isolation valve
target chamber

Fig. 3.3.4.1

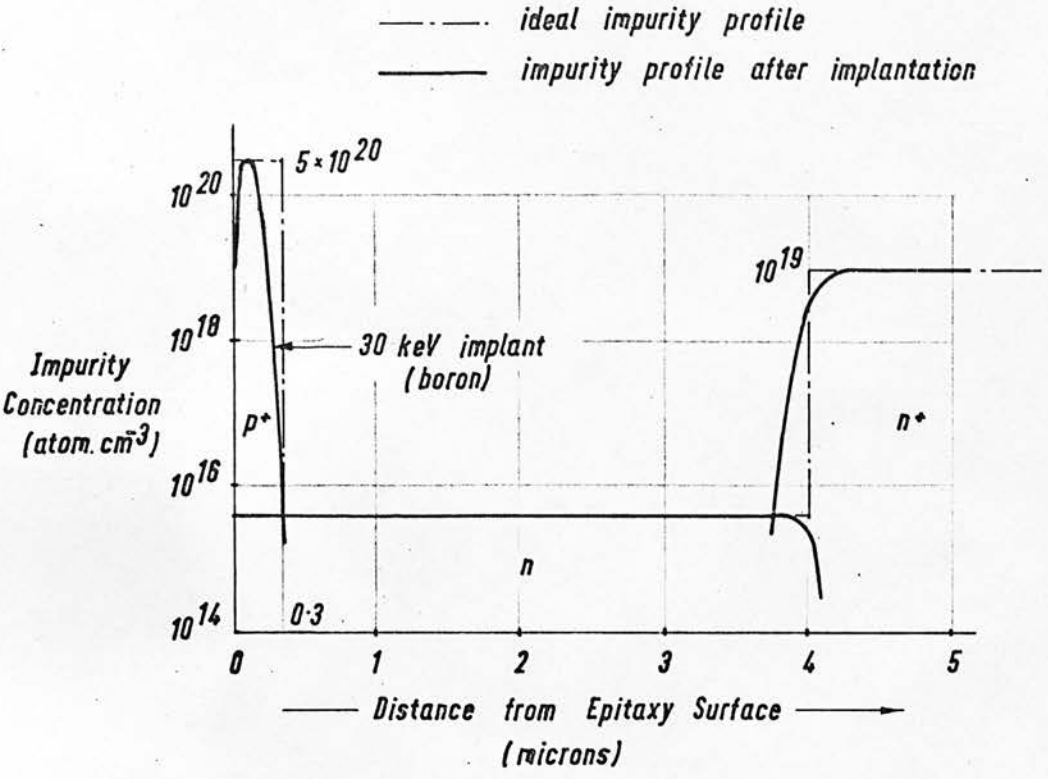
current monitoring, and thermal control. In consequence, steps must be taken to isolate source adjustment from the E.H.T. Supply. Vacuum requirements are met by use of oil diffusion and rotary backing pumps, providing the accelerator and target regions with a stable pressure $< 10^{-6}$ torr. An isolation valve is incorporated to allow target chamber access while the remainder of the system is under vacuum.

From the results of work carried out by Kleinfelder[39] on single energy implants, and Ferber[37] on implantation profiling, it is anticipated that a variable accelerator capable of 0-30 keV would enable the $p^+ nn^+$ impurity distribution shown in Fig. 3.3.4.2 for a (111) orientated single crystal silicon sample aligned to avoid channelling. With a beam current of approximately 10 μA., it would be possible to obtain an ion dosage of 3×10^{17} cm^{-2} per centimetre per hour, from which an estimation of total process time may be made.

3.4 ALLOY JUNCTION FORMATION IN EPITAXIAL SILICON

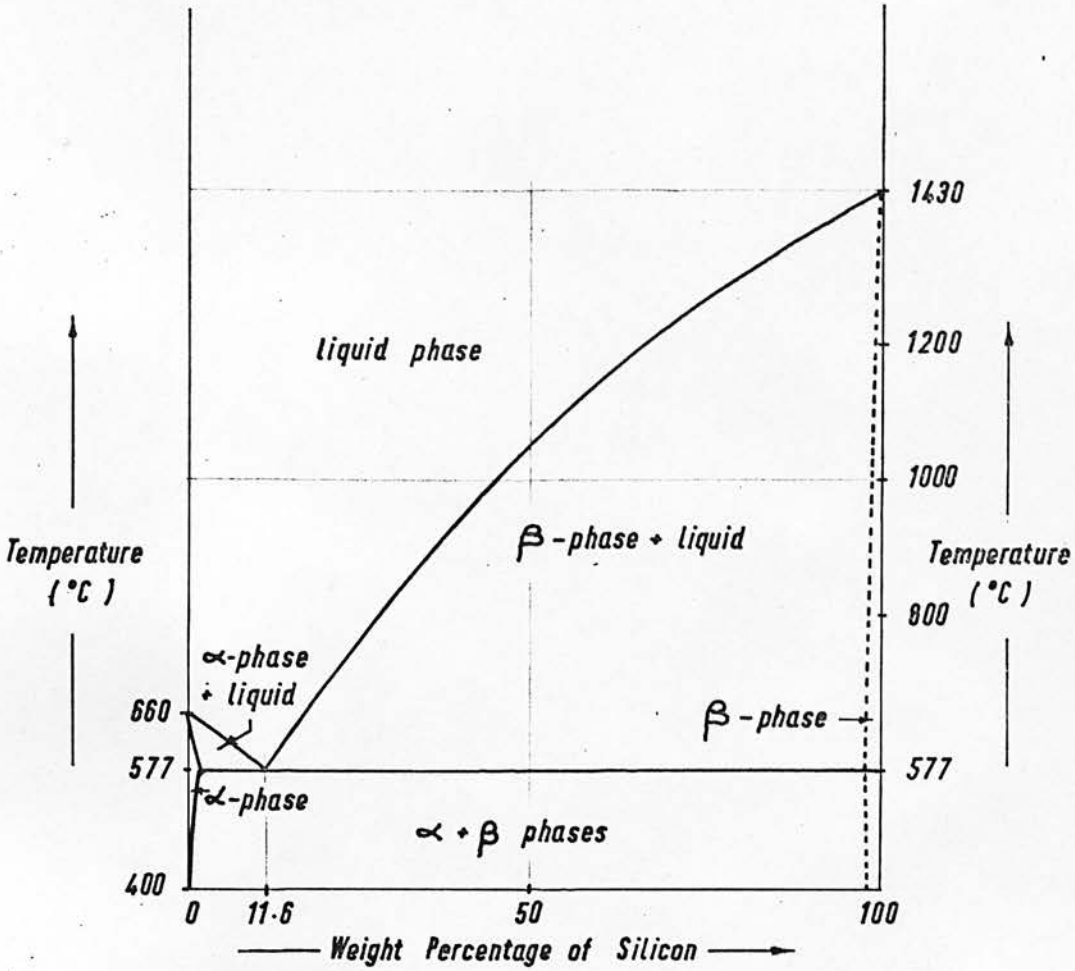
It was demonstrated by Pearson and Sawyer[40] in 1952 that, under certain experimental conditions, p-n junction diodes could be formed in an n-type silicon substrate by alloying aluminium into the substrate material. The mechanism of this impurity doping can best be explained by reference to the phase diagram for an aluminium-silicon combination.

Fig. 3.4.1 depicts the melt composition which is in equilibrium with a solid phase at any specific temperature.



Experimental Impurity Profile for Ion Implantation into Epitaxial Silicon

Fig. 3.3.4.2



Aluminium - Silicon Phase Diagram

Fig. 3.4.1

Although it is most easily interpreted for homogenous mixtures of elemental aluminium and silicon, it may be applied to a situation where the two materials, initially in contact, are subjected to a thermal cycle. When this combination is heated to a temperature in excess of 577°C and below the melting point of either constituent, the interface will liquify as both materials dissolve to meet dynamic equilibrium conditions. This situation creates two solid/liquid interfaces which separate as silicon from the substrate and aluminium from the overlayer are dissolved. In effect, a high concentration of aluminimum and silicon within the liquid at the respective interfaces will cause precipitation of α- and β- phases. As the process continues, and the silicon concentration increases towards the surface, the aluminium will be absorbed entirely giving rise to α- phase initially, then liquid phase, and finally β- phase silicon.

It would appear desirable to terminate the alloying process at an intermediate stage in order to reduce the fusion depth and retain regions of aluminium in α- and elemental forms which would enable good ohmic contact on the surface. During gradual cooling, the single crystal silicon substrate provides nucleation centres for the β- phase precipitation which forms an extension of the original crystal lattice doped with aluminium at a solid solubility level > 6×10^{18} atom. cm^{-3}.

In order to maintain a planar alloy-silicon interface, it is necessary to arrange the direction of alloy penetration to

be at right angles to the slowest dissolving crystal plane. This restricts substrate material to (111) orientation, and the junction profile is determined by time-temperature cycles together with the choice of original aluminium thickness. Since alloying is performed at temperatures significantly below those of conventional silicon processing, out-diffusion at the physical n-n^+ junction will be minimal and the ideal structure considered in Chapter 2 well approximated.

3.5 SELECTION OF A P^+ NN^+ SILICON DOPING PROCESS

Preceding sections in this chapter have described various methods of fabricating a p^+nn^+ structure in silicon. Final selection of an experimental process was made with the aim of achieveing greatest original development within a limited budget and project time scale.

The discussion of diffusion into single epitaxy revealed a need for more complex design analysis to accommodate distributed junction profiles. Furthermore, considerable experimental effort had already been expended on this particular type of construction, and an additional contribution would have been restricted by the time consumed in achieving a 'state of the art' condition within our laboratory.

Double epitaxy was given serious consideration as a technology still in its infancy and capable of providing the requisite abrupt junctions. Unfortunately, early plans to convert

an existing epitaxial reactor for this purpose were thwarted by the demands of other research to which the equipment was committed. Construction of a further reactor was to involve much duplication of effort, and the existence of two similar plants could not be justified readily. For these reasons, the process of double epitaxy was held in abeyance while alternative techniques were examined.

Ion implantation was regarded as an extremely useful method of achieving abrupt junctions in silicon; however a provisional cost estimate for purchasing even the simplest equipment appeared outwith available resources. The possibility of designing and building a modest system within our Department was given detailed consideration, from which it was decided that protracted equipment development of this nature would have a detrimental effect on the total research program.

Although alloy junction formation had fallen from favour within the semiconductor manufacturing industry, it did provide a low temperature method of junction formation, and required significantly less capital resources than any of the alternative processes already mentioned. Since the necessary equipment was readily available, and an experimental phase could be initiated rapidly, it was concluded that some investigation should be undertaken to demonstrate the applicability of this technique, with a view to proposing an inexpensive and efficient method of silicon impurity profiling for abrupt junction IMPATT diode construction.

3.6 DEVICE FABRICATION FROM A PROCESSED SILICON SLICE

Intimately connected with the method of silicon processing, are the problems of device isolation, separation, and mounting. In the manufacture of an IMPATT diode, these stages are extremely critical and largerly dictate the details of earlier processing.

Unfortunately, the need to obtain a uniformly abrupt junction, minimised electrical series resistance, and optimum thermal conductivity from the active region preclude many of the assembly techniques currently available to the semiconductor integrated circuit (S.I.C.) technologist. Because of the interdependence of these assembly stages, it is difficult to present a logical discussion of alternative solutions to each process stage chronologically. It is proposed therefore, to demonstrate the constraints which device assembly places on earlier silicon processing, and the compatibility of particular isolation techniques with subsequent separation, mounting, and ultimate requirements in device performance.

In conventional S.I.C. technology, devices are defined during selective diffusion doping stages by a technique of 'in contact' masking; while separation is performed in a 'scribe and break' operation as orthogonal grooves are cut along desired cleavage lines in the silicon and distributed force is applied to the entire slice to initiate 'chip' detachment.

This approach has the disadvantage that subsequent device assembly is not readily amenable to mechanical automation.

Processed chips must be individually located, transferred to a mounting base, and physically secured prior to wire bonding. More important, however, is the restriction imposed by physical attachment: an alloying operation, which demands that devices are assembled with the substrate material adjacent to the mounting base, otherwise silicon utilised in fusion will disturb an active region close to the processed surface. By mounting the diode accordingly, thermal conductance from the junction is restricted by the presence of undesirable substrate material inhibiting heat transfer to the heat sink. Substrate thinning can be employed to minimise this effect, but the resultant fragile nature of an integral slice renders device separation difficult, and alloy mounting becomes a hazardous operation.

The alternative configuration of device assembly, where the junction side is in contact with the mount, has certain attendant problems. Any such arrangement is incompatible with an earlier process which defines a localised p^+ region having a surface coplanar with and surrounded by an n^- region, since a short circuit condition will be introduced at the device-mount interface. This situation can be avoided by mesa etching, where undesirable n-type material is physically removed from the periphery of the heavily doped p-region; but selective etching demands a photolithographic stage, where mask registration to a previously alloyed area poses an additional complication. For this reason, it was decided that alloying should be performed over the entire slice, and the problem of mask registration eliminated

by defining the active device area in a post-mounting operation.

The choice of a chip assembly technique which would provide the required physical, thermal, and electrical properties without utilisation or even damage of the silicon crystal was limited to epoxy adhesive and thermocompression bonding.

Use of silver or gold loaded epoxy adhesive allows direct attachment of a silicon chip to the mount, resin curing can be accomplished at temperatures below 200°C, and the nature of the bond permits a wide range of mounting materials. While this low temperature technique requires minimal equipment and little sample preparation, it suffers from inferior thermal, electrical, and chemical properties of the epoxy layer.

By comparison, thermocompression bonding, previously associated with gold wire bonding to aluminium or gold lands, provides more satisfactory physical properties; however sample modification is necessary since the process depends on the fusion of two metallic surfaces subjected to heat and a positive pressure contact. It was anticipated that preparatory gold deposition on the silicon and mounting surfaces would enable a bond which was resistant to any later chemical processing.

Device definition within a mounted chip can be achieved by chemically etching until complete substrate penetration has taken place. This prolonged operation suffers from two severe problems however. The masking properties of photoresists

deteriorate with exposure to the acids involved, and lateral etching will cause significant deviation of physical profile from that defined by the original mask. While the latter problem of under-etch may be compensated for in the mask design, it is impractical to accurately control chemical etching of a device 100 μm in diameter while subjecting the remaining chip to a vertical stock removal of 300 μm.

An alternative approach incorporates an ultrasonic drilling technique used by semiconductor device manufacturers in the production of large area diodes. This process involves the erosion of unwanted silicon material using a fine abrasive slurry activated by the transmission of ultrasonic energy from the end of a drilling tool. A critical positive pressure is applied to the tool, thereby loading the sample and enabling tool penetration as erosion proceeds. Since the pattern imparted to the silicon will be the inverse of that on the drill, cylindrical devices may be fabricated using a tool which has an array of holes machined parallel to the direction of intended abrasion. These recesses will allow the requisite passage of silicon while the surplus is removed.

While it was appreciated that ultrasonic drilling would alleviate the need for a photolithographic facility and exhibit negligible lateral erosion compared with chemical etching performance, the practicability of machining devices an order of magnitude smaller than typically manufactured was not certain.

It was imperative therefore, that any commitment to this technique be preceded by an experimental feasibility study.

Device assembly has already been shown to be a critical operation in the fabrication of an IMPATT diode. Microwave output is dependent on the effectiveness of electrical and thermal conduction from the active region, and these properties are functions of device orientation within the mounting package, the intimacy of bond between device and package, and the material used as a mounting base. The provision of a composite encapsulation which would permit satisfactory electrical performance requires considerable development and tooling effort. For the purposes of laboratory demonstration such work was not of primary importance, and thought was directed towards simple mounting which would permit functional testing of diodes in an unencapsulated form. The chosen construction will be discussed in Chapter 4.

3.7 THE CONCEPT OF INTEGRAL SLICE PROCESSING

Section 3.6 has indicated a range of possible techniques which could be used at various stages of device assembly. In planning an experimental program however, it was necessary to collate every aspect of device fabrication and select process methods which would demonstrate compatibility throughout. Although single devices could be made by permutation of method until an optimum combination was achieved, such a solution was rejected because of likely device yield and fabrication time which could be criticised in sequential processing applied to individual

diodes.

In view of these shortcomings, some time was devoted to an examination of the potential implications of successful integral slice mounting. This operation, which appeared attractive in reducing bonding effort and eliminating the need for dexterous chip assembly, could be followed by device definition, isolation, and separation stages, to yield individual components for immediate electrical testing. A batch manufacture of this nature has considerable advantages during the diagnosis of early device failures, and must ultimately be an asset in reducing process time.

Nevertheless, several process difficulties arose directly from this approach, and they had to be assessed before final commitment to an experimental schedule. Inversion of mounting and device separation in the assembly sequence renders a scribe and break operation ineffective, and accordingly chemical or ultrasonic abrasive etching must be used to define and isolate the diodes. Furthermore, since epoxy adhesive cannot withstand chemical etching treatment, it is necessary to concentrate on a thermocompression silicon attachment process as a precaution against failure of abrasive device isolation.

Chemical resistance is also required of the mounting material, and an obvious protection consists of complete superficial gold plating.

While the problem of final diode separation from a common

heat sink can be tackled by precision machining, it was hoped that the ultrasonic drilling technique developed for diode separation could be adapted for this purpose by the simple expedient of changing the cutting tool.

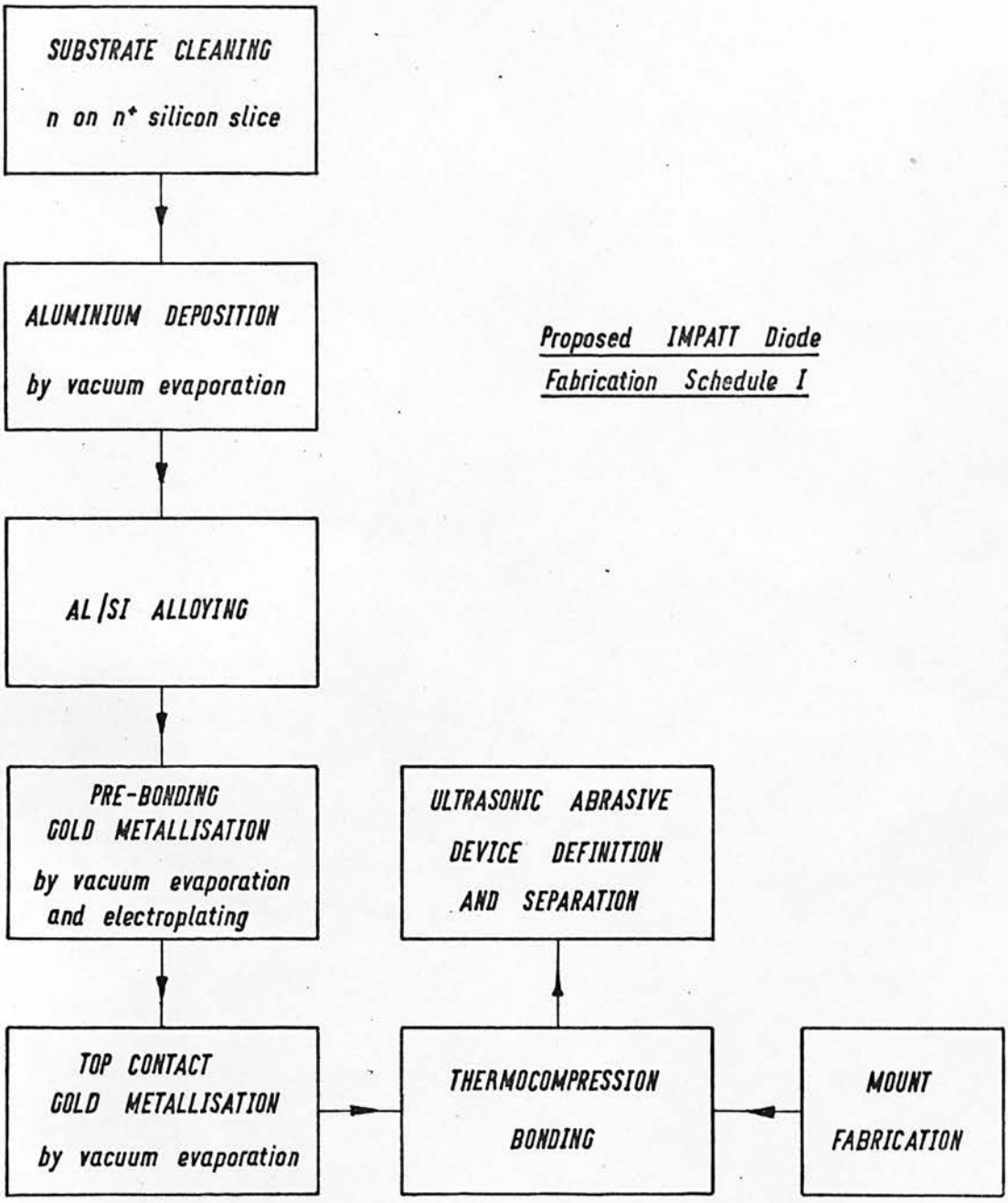
3.8 AN EXPERIMENTAL PROCESSING PROGRAM

While the foundations of all the constituent techniques required for integral matrix formation are in existence, few have been developed in the manner required for this work, and none has been sufficiently well documented for immediate reproduction. Contrasted with conventional technology having a known capability of device yield early in an experimental program, this alternative must appear dubious, depending on the success of a sequence of uncertain processes.

In the knowledge that process failure might inhibit ultimate success in integral matrix formation, a choice had to be made between the development of an original diode fabrication technique and a reappraisal of research into IMPATT diode assembly using conventional technology. It was considered that little benefit would accrue from duplicating effort within a well established process and that insufficient contribution could be made to the improvement of a device which had already developed for many years in this environment. The more ambitious program might not succeed in yielding superior devices with improved geometry and reduced fabrication cost, but would certainly provide valuable information about a new assembly technique which could

have more general implications in the future of semiconductor device manufacture.

In conclusion, the experimental program was directed towards development of alloy junction formation and integral slice processing, in the hope that the schedule indicated pictorially in Fig. 3.8 would ultimately enable production of avalanche diodes capable of microwave output.



Proposed IMPATT Diode
Fabrication Schedule I

Fig. 3.8

CHAPTER 4

EXPERIMENTAL DEVELOPMENT OF AN IMPATT DIODE FABRICATION SCHEDULE

In view of the diversity of technology involved in the various stages of IMPATT diode fabrication as described in the last chapter, it was decided that individual process problems would be considered in isolation prior to a systematic combination which would ultimately yield a complete device.

At an early stage in the experimental preparation, it was established that n on n^+ silicon slices processed to our specification were readily available from research laboratories investigating low temperature epitaxial growth. In consequence, this aspect of fabrication was ignored, and development effort was directed towards the production of a satisfactory alloying technique, the provision of an integral slice thermocompression mounting facility, an investigation into methods of device isolation and separation, and the establishment of test equipment for final device appraisal.

4.1 AN INVESTIGATION OF P^+N ALLOY JUNCTION FORMATION

Because of the exacting junction requirements of an IMPATT diode, preliminary investigations had to be carried out to establish a satisfactory aluminium deposition technique and a controlled alloying process which would enable abrupt and uniform junction formation at a predictable depth.

It was necessary to consider the possibility of deriving a theoretical analysis to describe the required reaction. Prediction of correct processing temperature and time parameters presents a formidable problem involving chemical reaction kinetics, and can only be interpreted using equipment capable of accurate and rapid temperature cycling. Since the available facility was not of such sophistication, it was decided that time would be more profitably spent on empirical rather than theoretical assessment of junction formation.

The experimental procedure adopted was to deposit a standard 2 μm thickness of aluminium on a chemically cleaned 1 Ω-cm n type silicon slice within a vacuum evaporator; to use photolithographic and etching stages to define a matrix of metallic dots 250 μm - 50 μm in diameter; to alloy these dots, deposit a gold ohmic contact on the n^+ surface; and to assess the resultant product in terms of physical and electrical parameters as a function of the time-temperature alloying cycle.

There are unique properties pertaining to such hemispherical junctions formed on high resistivity silicon, which would not appear in the final device: nevertheless, it was reasoned that data obtained during this preliminary work would be indicative of process requirements during uniform alloy formation in n on n^+ substrates.

4.1.1 SLICE PREPARATION

The/

The original schedule used in silicon slice preparation is detailed in Appendix D. By virtue of the potential hazards associated with such a process, it was necessary to confine slice cleaning to a fume cabinet located some distance from the vacuum evaporator used for aluminium deposition. During initial laboratory work however, difficulty was encountered in obtaining satisfactory aluminium adhesion to the silicon surface, and this was attributed in part to substrate contamination arising from excessive exposure to normal atmospheric impurities. By transferring slices from the cleaning cabinet to the vacuum system location under iso-propyl-alcohol, and in addition, providing a substrate heater within the chamber to enable 'bake out' and subsequent deposition at an elevated temperature, the problem was eliminated.

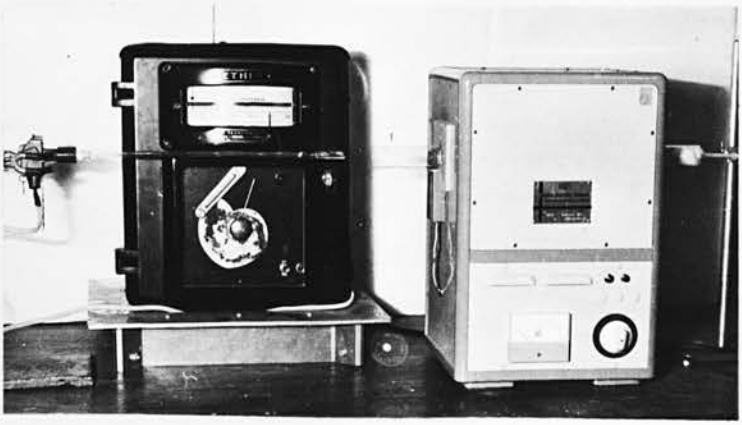
At a later stage in electrical testing of alloyed junctions, it became apparent that the cleaning technique was still inadequate and an alternative schedule, detailed in Appendix E, was substituted. Proving effective and comparatively innocuous in operation, it was used for all subsequent device work.

4.1.2 ALUMINIUM DEPOSITION

Aluminium deposition was performed in a standard Edwards High Vacuum 19E vacuum system at a pressure $< 10^{-5}$ torr obtained by the use of oil diffusion and rotary backing pumps. The work space, shown in Plate 4.1.2., contained a six point rotary filament holder which could be equipped with tungsten spiral sources to

Vacuum System Work Space

Plate 4.1.2



Alloying Furnace

Plate 4.1.4

permit upward evaporation of several aluminium charges during one evacuation cycle.

Single samples were located 12 cm. above the operative source, and maintained at 100°C during evaporation by means of a resistance heater placed in intimate contact with the reverse surface of the silicon slice. Temperature control was effected from a thermocouple sensor within the heater block, which was coupled to a Eurotherm controller regulating the electrical supply.

A predetermined mass of 5-N aluminium was evaporated in entirety in order to obtain the requisite overlayer thickness, and subsequent confirmation achieved by direct measurement using a Talysurf.

Early difficulties with the alloying cycle were traced to organic contamination introduced by the pumping system during aluminium deposition: accordingly, it was necessary to introduce a liquid nitrogen fed cold trap immediately above the baffle valve, which condensed vapour escaping from the diffusion pump.

From this modification, a standard deposition sequence was developed which provided aluminium films with good adhesion and adequate purity over a thickness range 0.1 μm - 5 μm . The schedule, detailed in Appendix F, was used exclusively during device fabrication.

4.1.3 PRODUCTION OF AN OVERLAYER PATTERN

Conventional/

Conventional photolithographic and chemical etching processes were used to obtain the requisite aluminium pattern. A matrix of discs having selected diameters within the range 1-5 cm was prepared on a rubylith master, and reduced by a two stage photographic process incorporating 'step and repeat' to provide a regular array of opaque dots 50-250 μm in diameter on a Kodak maximum resolution plate. Shipley AZ 1350 photoresist was liberally applied to the silicon sample through a hypodermic syringe fitted with 1 micron filtering, and the surplus removed during a spinning operation. After a drying period, the sample was subjected to ultra violet exposure in the regions defined by the interposed photographic plate, and subsequent development followed by low temperature baking prepared the aluminium layer for chemical etching. It was found that immersion in a solution of

46 g. ammonium persulphate
+ 54 ml. de-ionised water
+ 50 ml. orthophosphoric acid (98% strength)

heated to 40°C and provided with ultrasonic agitation resulted in optimum resolution of the desired pattern. Protective photoresist was removed with a proprietary solvent and the sample given a rinse in boiling iso-propyl-alcohol before final drying in a nitrogen atmosphere. This completed slice preparation prior to an alloying cycle.

4.1.4 THE EXPERIMENTAL ALLOYING FACILITY

The apparatus used for alloying is shown in Plate 4.1.4.

It consisted of a Griffin and George 1.25 in. bore muffle furnace fitted with a quartz liner; a quartz paddle which supported a sample within the furnace; and a temperature controller, coupled to a thermocouple inserted in the centre of the paddle, to regulate the supply of power to the heater. In addition, a continuous flow of gaseous nitrogen was passed through the furnace tube at a rate of 500 ml. min^{-1}.

Although this equipment did not permit high accuracy sample temperature measurement, it was possible to obtain good reproducibility of experimental conditions during successive alloying operations involving simple thermal cycles. Temperature gradients across the sample were observed, and taken into account in result interpretation.

Rapid thermal cycling could only be achieved by manual adjustment of sample position within the preheated furnace. This technique gave inconsistent results between experiments, and consequently, efforts were made to develop a process which did not depend on such manipulation.

4.1.5 ALLOYING EVALUATION

Interpretation of alloy formation depended on physical examination and electrical testing.

One method of junction depth measurement involved the 'angle lap and stain' technique explained at length in Burger and Donovan[41]. Samples were lapped at a shallow angle (1°-5°) on a

glass plate, using 0.1 μm diamond abrasive with a water soluble lubricant. Subsequent cleaning and staining were performed to reveal regions of different impurity type. Although adequate p-n delineation could be achieved by 30 second exposure of the sample to a solution of 0.1% concentrated nitric acid in hydrofluoric acid (48%) under strong light, the alternative stain

2 g. $HIO_4 . 2H_2O$
+ 1 ml. HF (48%)
+ 20 ml. H_2O

suggested by Nicolau[42] was preferred, since it had the additional property of indicating significant changes in impurity level within either p or n regions. This was considered advantageous for later work involving n on n^+ substrate material.

The measurement of junction depth was finally carried out by microscopic interferometry using a sodium light source, and by estimating to the nearest quarter fringe (one eighth wavelength) an accuracy of $\pm$ 0.08 μm could be obtained.

An alternative technique made use of the differential chemical etching property observed by Calverley[43] et. al. to exist due to the built-in field at p-n junctions. This modification to the simple profile of a mesa structure enabled extremely accurate determination of abrupt changes in impurity level by direct measurement on scanning electron micrographs. Mesas were formed in alloyed samples using a black Apiezon wax mask to restrict the

silicon etch,

3 parts acetic acid
+ 5 parts nitric acid
+ 3 parts hydrofluoric acid
by volume of analar chemicals.

The effectiveness of this process will be demonstrated in Section 5.5 for a complete $p^+n\,n^+$ structure.

The quality of an alloyed junction was assessed by examination of diode d.c. characteristics. Lindmayer and Wrigley[44] explain this process in an account of semiconductor diode operation, where they describe the conditions existing in a narrow depletion region, and the theoretical voltage-current relationship for forward and reverse biasing. This familiar and simple analysis provides an equation for diode current *(I)*

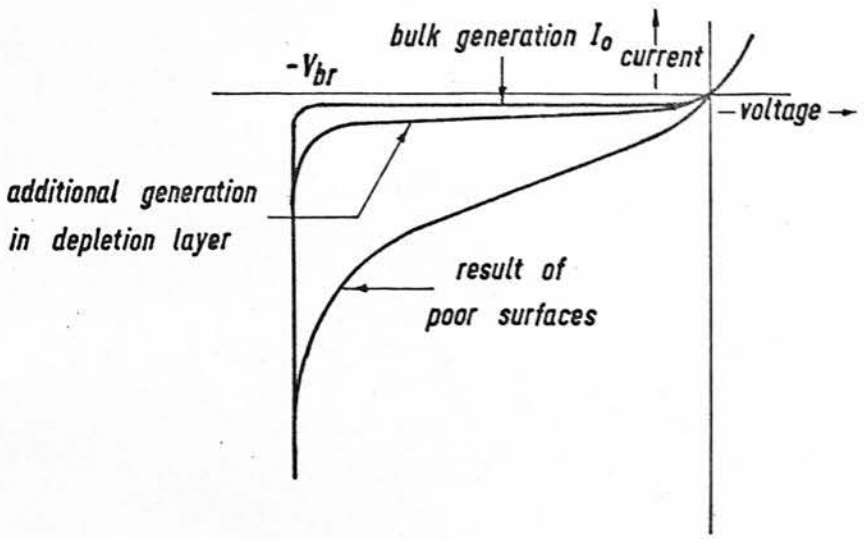
$$I = I_0\left(e^{\frac{qV}{kT}} - 1\right) \qquad 4.1.5.1$$

where I_0 is the reverse saturation current, and V is the applied voltage. The mechanism of current flow which gives rise to a saturation value I_0 for reverse bias is thermal carrier generation external to the space charge region and initiated in an attempt to reinstate equilibrium disturbed by the depletion of carriers from the junction. Because of the increasing depletion width with reverse bias, account must also be taken of complementary internal thermal generation. The magnitude of this additional current is obviously dependent on the degree

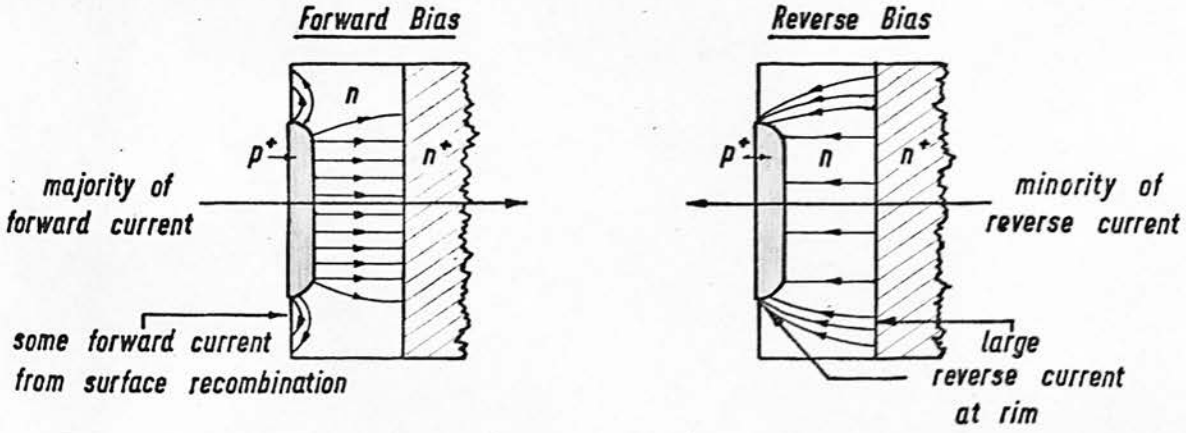
of depletion, or according to equation 2.2.1, is proportional to $\sqrt{V}$. As the voltage is increased towards a breakdown level, multiplication amplifies this current and modifies the reverse characteristic from the simple form as shown in Fig. 4.1.5a.

In real terms however, leakage currents are observed to be considerably higher than predicted by such theory. The discrepancy arises from surface imperfections at the boundary of the junction which reduce carrier lifetime. In the absence of silicon crystal defects, the origin of these imperfections lies at the substrate crystal surface which is susceptible to contamination despite the cleaning procedure adopted. Absorption of many foreign elements will occur in gaseous or vapour form from the environment, and in precipitate from the cleaning process. An additional impure aluminium deposition stage will also contribute, since subsequent junction formation may introduce contaminants providing recombination centres which will reduce carrier lifetime and promote surface currents.

This situation is depicted in Fig. 4.1.5b reproduced from Lindmayer and Ridley[44]. When the junction is forward biased, increased surface recombination rate will not significantly affect total current flow, since the ohmic contact is still the most appropriate sink for hole current. It will be seen however, that application of reverse bias creates a condition where short carrier lifetime at the surfaces allows current flow at the junction rim, thus enhancing the level of thermal generation.



(a) Reverse Bias Diode Characteristics



(b) The Effect of Surface Recombination on $P^+N N^+$ Diode Operation

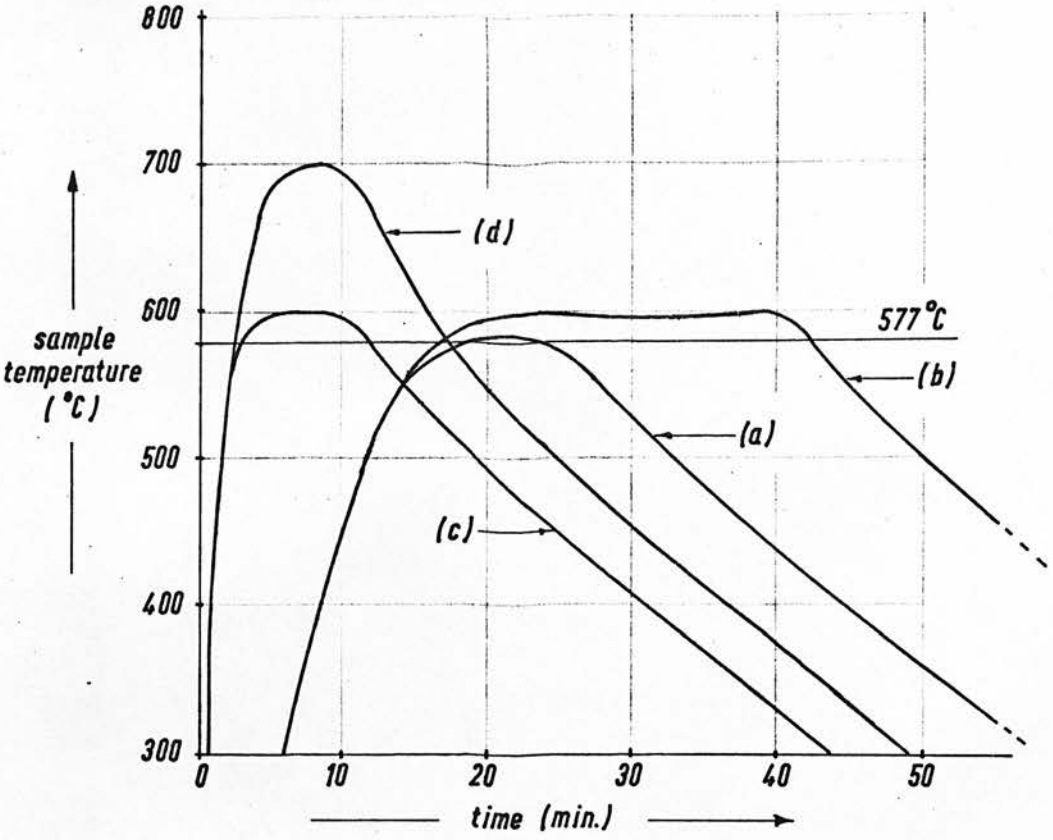
Fig. 4.1.5

The degree to which such an effect occurred in experimental structures was used as a measure of alloying purity and uniformity.

4.1.6 OPTIMISATION OF ALLOYING

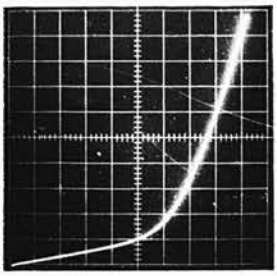
It was anticipated that control over alloy formation could best be achieved at low temperatures, since the reaction rate would be minimised. Early experiments were conducted using a maximum temperature of 580°C, according to the thermal cycle shown in graph (a) of Fig. 4.1.6. Samples subjected to this treatment showed no physical evidence of junction formation when scrutinised by the lap and stain technique, and the aluminium layer appeared unaffected by processing. Matrix 2-70 was studied under d.c. electrical test from which typical forward and reverse biased junction characteristics are shown in Plate 4.1.6 (a), (b). In the light of these results it was considered probable that slight alloy formation had occurred in isolated regions of the silicon-aluminium interface. Electrical performance was attributed to poor order of α- and β- phases within the reaction zones, and more significantly, the parallel effect of an aluminium silicon contact which had been unaffected by processing.

It was concluded that a satisfactory reaction could only be initiated by raising the process temperature: accordingly matrix 4-70 was subjected to the thermal cycle of Fig. 4.1.6 (b) with an upper temperature limit of 600°C. Plate 4.1.6 (c) (d) demonstrates that resultant junctions exhibited the theoretically



Thermal Cycles for Al/Si Alloying

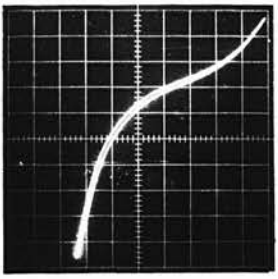
Fig. 4.1.6



a. Fwd. bias

scale : horiz. = 0·1V./div.

vert. = 1mA./div.

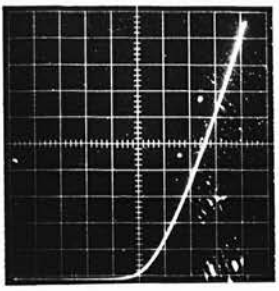


Alloying Temperature = 580 °C

b. Rev. bias

scale : horiz. = 10V./div.

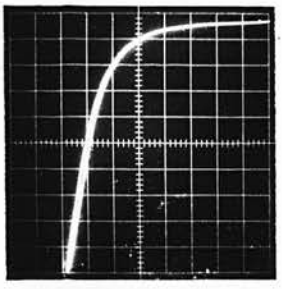
vert. = 1mA./div.



c. Fwd. bias

scale : horiz. = 0·1V./div.

vert. = 1 mA./div.



Alloying Temperature = 600 °C

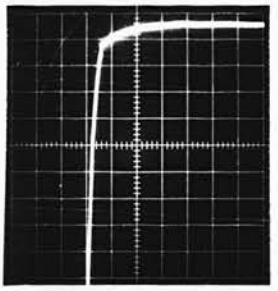
d. Rev. bias

scale : horiz. = 10 V./div.

vert = 1 mA./div.

Alloyed Junction Characteristics

Plate 4.1.6(a - d)

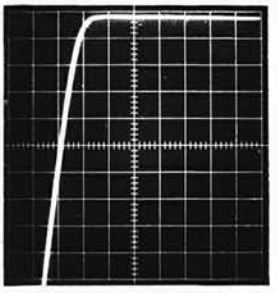


Alloying Temperature = 600 °C

e. Rev. bias

scale: horiz. = 10V./div.

vert = 1mA./div.

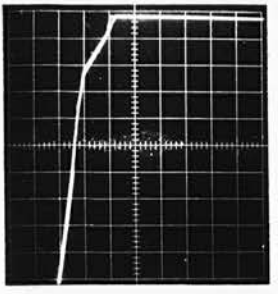


f. Rev. bias

scale: horiz. = 10V./div.

vert. = 1mA./div.

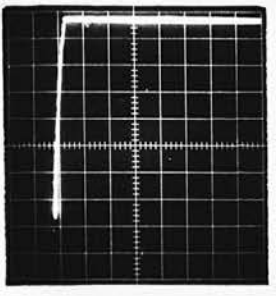
Alloying Temperature = 620 °C



g. Rev. bias

scale: horiz. = 10V./div.

vert = 0·01mA./div.



Alloying Temperature = 700 °C

h. Rev. bias

scale: horiz. = 10V./div.

vert. = 1mA./div.

Alloyed Junction Characteristics

Plate 4.1.6 (e - h)

predicted forward bias characteristic with a slope resistance corresponding to the high resistivity substrate material outwith the active zone. Reverse bias performance was very unsatisfactory however, indicating premature breakdown, and excessive leakage current prior to avalanche. Physical measurements revealed uneven alloy penetration approaching 4μm in places, and it was deduced that thermal cycling would have to be modified in order to reduce the time at elevated temperature.

Since Matlow and Ralph[45] had already established that slow cooling was imperative to epitaxial regrowth, a method of rapid heating had to be incorporated to reduce the total reaction time. By preheating the furnace to the maximum desired temperature before sample insertion, it was possible to modify thermal cycling as indicated in Fig. 4.1.6 (c). This alteration resulted in an improved physical appearance of processed samples (matrix 7-70) which exhibited uniform alloying to a depth of 1.2μm . While the reverse breakdown characteristics also improved, Plate 4.1.6 (e), the level of leakage current at high reverse voltages was still excessive, and the avalanche threshold detectably lower than predicted.

A further increase in alloying temperature to 620°C was implemented and the desired reverse breakdown condition of Plate 4.1.6 (f) finally obtained from matrix 15-70.

At this stage, the electrical characteristics were examined in greater detail, and it was established that the

maximum dissipation level for all diodes was similar, being independent of junction area. Furthermore, at low avalanche currents, distributed breakdown was observed as reproduced in Plate 4.1.6 (g). These results were symptomatic of the filamentary (microplasma) breakdown mechanism reviewed by Kressel[45].

Attempts to obtain greater junction uniformity by further varying the time-temperature processing parameters in the eutectic region were in vain. This suggested that alloying had been initiated at isolated locations throughout the aluminium silicon interface, and that further reaction arose mainly from these regions of liquid phase. In such a situation, penetration would be governed by an advancing liquid surface and vary in depth to a maximum under the centres of initial alloy formation. It was decided to investigate this theory by alloying at temperatures above the melting point of aluminium, thereby introducing a liquid phase early in the thermal cycle which might enable more uniform reaction at the interface of the two materials.

Experiments were conducted incorporating the diagnostic techniques already mentioned to derive a suitable thermal cycle which was found to be more critical as a result of the increased process temperature. A schedule was eventually developed which gave rise to junction formation at a depth of 2 μm. Sharp junction breakdown was observed at 80 volts, and the microplasma effect limited to a minority of alloyed areas in any sample. The

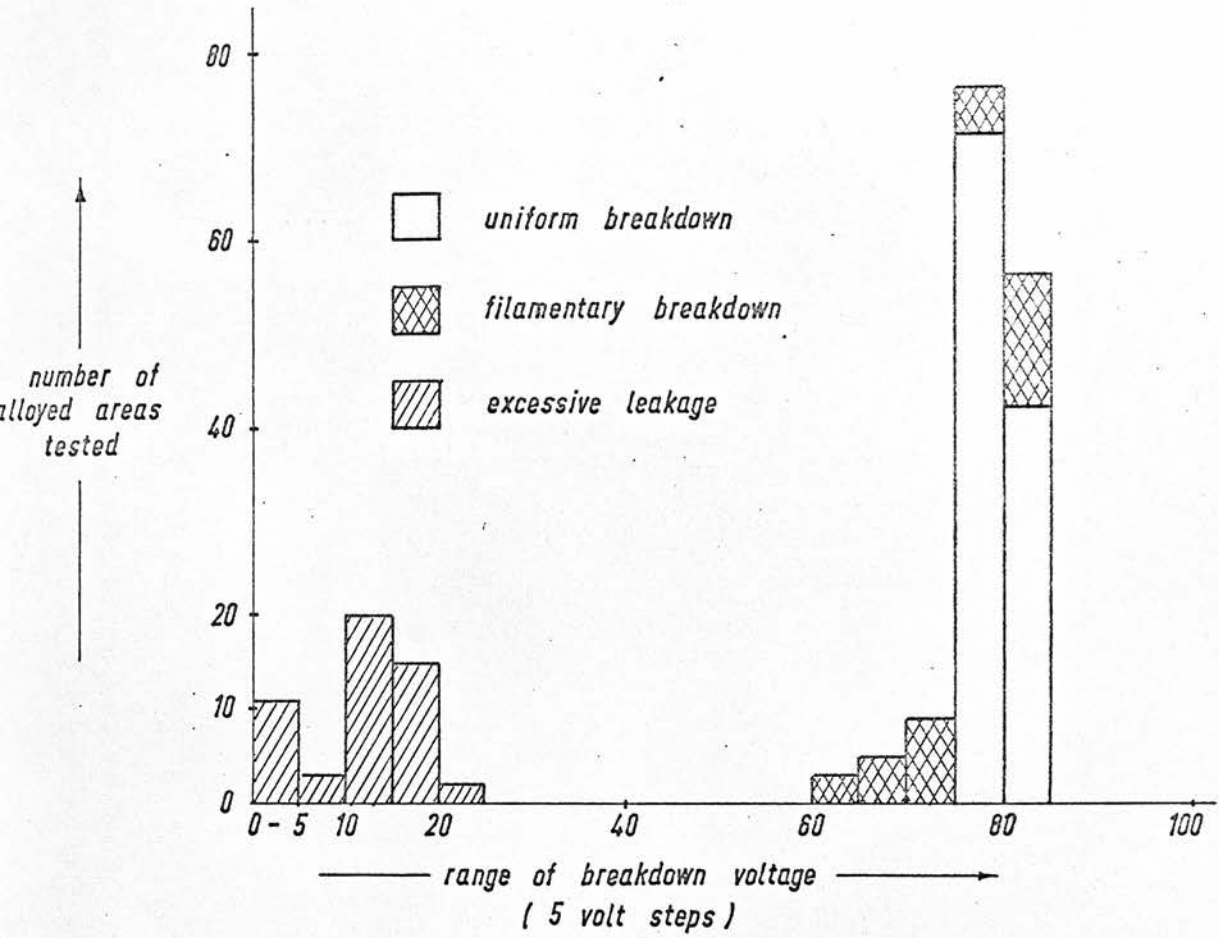
thermal cycle shown in Fig. 4.1.6 (d) was applied to matrix 22-80 from which the low current breakdown characteristic for 100μm diameter alloyed region is reproduced in Plate 4.1.6 (h). This d.c. performance was considered satisfactory, and further electrical measurements were devoted to an appraisal of yield and reproducibility.

Three additional slices were similarly processed and approximately 200 alloyed areas subjected to evaluation. Interest was confined to the central region (1cm. in diameter) of each slice, since microscopic observation revealed irregular aluminium utilisation around the perimeter : this was attributed to temperature gradients within the furnace.

Results of this investigation are summarised in the histogram of Fig. 4.1.7 where distinction is made between three electrically different types of alloyed junction exhibiting

1.	Filamentary breakdown	60 - 85 volts	(15%)
2.	Excessive surface leakage	>1mA. at 80 volts	(25%)
3.	Uniform breakdown	78 - 81 volts	(60%)

The spread in breakdown voltage from devices in the third and satisfactory category was derived from the use of four separate slices, indicating a variation in substrate resistivity of less than 3 percent. Variations in breakdown voltage on any given slice were, to a volt measurement accuracy, imperceptible.



Histogram of Alloying Performance
(schedule 'd' fig. 4.1.6)

Fig. 4.1.7

This alloying yield provided a satisfactory basis for experimental device work, and effort was directed towards later stages of diode fabrication.

4.2. INTEGRAL SLICE THERMOCOMPRESSION MOUNTING

Although the success of batch diode formation depended on several consecutive operations, there was only one process for which there could be no alternative. Since an inability to perform integral slice thermocompression mounting would prevent further fabrication, it was decided that time should be devoted to this aspect early in the experimental phase.

In 1968, Hambleton[47] et al. announced their success in thermocompression bonding a large area of silicon to a gold plated copper mount. This achievement involved thinning the silicon slice to 10 μm $\pm$ 1 μm, lapping the copper mount to obtain a combined parallelism of better than 1 μm over the sample area, and coating both mating surfaces with a 1 μm thick layer of gold prior to bonding at 150°C with a pressure of $3.5 \times 10^{7} kg.m^{-2}$. This information provided a reassuring guide for our endeavour; however the description of equipment, process time, and sample area were insufficient to allow immediate replication of Hambleton's accomplishment.

Our laboratory experience of thermocompression attachment

was associated with gold wire ball bonding to metallic films, where contact areas of approximately $1.5 \times 10^{-8} m^2$ were fused at a temperature of 300°C and a pressure of 10^6 kg.m^{-2}. In view of the intention to perform integral slice bonding at reduced temperature, it appeared relevant to establish the thermal dependence of bonding pressure for this miniature system. There were more serious uncertainties however, connected with physical sample response to the forces involved, which might ultimately have prevented successful bonding, or at least restricted maximum sample area and the total force applied. Machining precision in compression jig and heat sink fabrication were ultimately to determine the degree of force uniformity and hence localised stresses over the sample area during bonding. The extent of this effect could not readily be predicted and consequently, priority was given to the design of a suitable thermocompression jig which would allow experimental observation of sample deformation arising from process conditions. A subsequent bonding development program could then be undertaken in the knowledge of apparatus limitations.

4.2.1 THE BONDING ALIGNMENT JIG

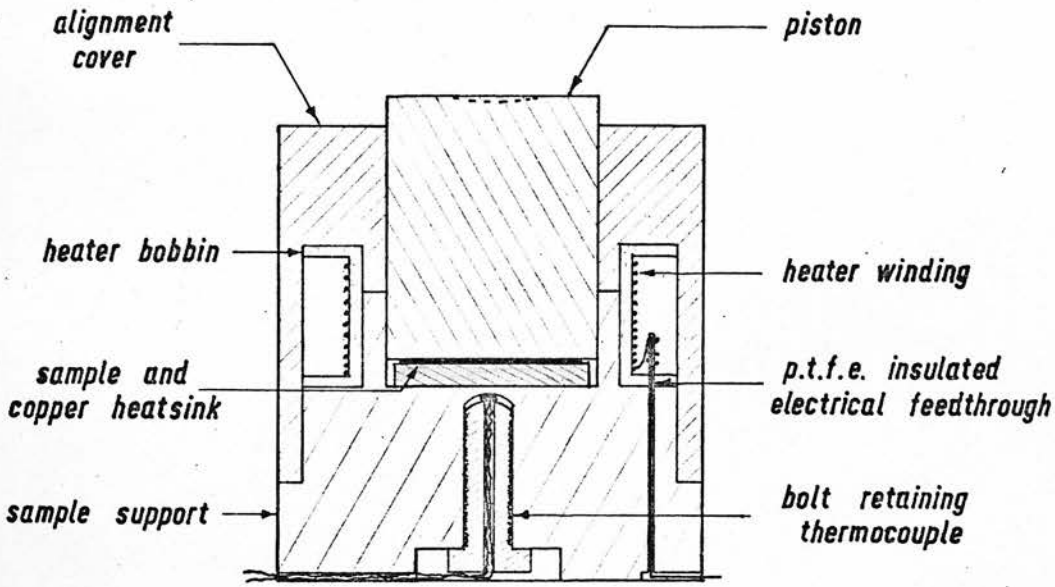
Design of an assembly jig for thermocompression mounting was based on the use of 2.5 cm diameter substrates, and distortion free operation under a maximum and uniform load of 10^4 kg., dictated by the capacity of available equipment. Provision was required for sample heating to 200°C with an appropriate monitoring facility.

The final construction is shown schematically in Fig. 4.2.1 which also indicates sample and heat sink location during bonding. With the exception of the piston which was formed from ground silver steel rod, all machined parts were constructed from mild steel. Extreme care was exercised in boring the central guide to ensure a perpendicular base and a piston clearance of $< 5\,\mu m$. Two features of the piston not indicated in the diagram are a vertical groove cut along its length to allow free passage of air during operation, and the work-face which was lapped with $3\,\mu m$ diamond abrasive to ensure planarity better than $2\,\mu m$. The heating element was wound from anodised aluminium wire and attached to P.T.F.E. insulated copper feedthroughs, while a chromel-alumel thermocouple was mounted in the vicinity of the sample for process temperature measurement.

4.2.2. SELECTION AND PREPARATION OF A HEATSINK

Although Swan[48] had indicated the importance of providing a good heat sink for IMPATT diodes by demonstrating improved performance of devices mounted on type II A diamond, the obvious choice for our work was copper, since in addition to having good thermal and electrical properties, it could easily be machined to meet final circuit assembly requirements. Furthermore, confidence could be placed in a material which had provided satisfactory results in a previous and similar application.

Discs of 2.7 cm diameter and 0.31 cm thickness were machined from a rod of high conductivity hard drawn copper, and



Bonding Alignment Jig

Fig. 4.2.1

subjected to a lapping procedure in order to improve planarity and reduce surface imperfections. Successively finer grades of diamond abrasive with water soluble lubricant were used on a cast iron lap to remove undesirable stock material, while a concluding polish on microcloth impregnated with 1μmdiamond gave rise to a sample surface exhibiting maximum irregularities < $\pm$0.1 μm and overall distortion < 1μm .

The next stage in preparation was to coat the heat sink with an overlayer of gold. It was considered most appropriate to use an electroplating solution for this purpose, since a thickness greater than 1μm was required to accommodate the remaining error in sample and mount surface parallelism, and because complete coverage was essential to protect the copper from any chemical treatment which might later be applied to the sample/heat sink combination.

After a cleaning period in trichlorethylene, copper discs were immersed in a Johnson-Matthey gold potassium cyanide plating solution which was thermostatically controlled about 60°C and lightly agitated with a magnetic stirrer. Platinum was used as the anodic electrode and a current density of 3 mA.cm^{-2} gave rise to an overlayer thickness of 5μm in 25 minutes. Subsequent immersion in boiling water (double distilled, de-ionised) for a short period removed cyanide contaminants from the plated surfaces, and drying in a nitrogen atmosphere completed preliminary heat sink treatment.

Initially unsuccessful thermocompression bonding attempts revealed severe discoloration of the gold plating, and this was traced to lapping contaminants which had become lodged below the surface of the malleable copper discs and had caused oxidation during the heating cycle. Accordingly it was necessary to modify the preparation schedule before electroplating to include a chemical etching stage which removed copper superficially to a uniform depth of 10 μm. The solvent in this case consisted of 5% hydrogen peroxide in ammonium hydroxide by volume of analar chemicals.

4.2.3 SAMPLE PREPARATION

Since the aluminium surface of the alloyed sample was not directly amenable to gold electroplating, it proved necessary to establish an initial film from which growth of a uniform and tenacious layer could occur. This was effected by twofold evaporation in the Edwards High Vacuum system during which nichrome and gold were sequentially deposited on the sample. A system pressure of $<10^{-5}$ torr was found to be adequate, and the silicon sample maintained at 150°C throughout. Nichrome, used to promote gold adhesion, was evaporated from a tungsten spiral source to provide a deposition thickness of 300 Å while continuous monitoring was performed by calibrated resistance measurement on a glass substrate positioned in the vapour stream. Subsequent gold layer deposition to a thickness in excess of 1500 Å involved a similar evaporation procedure from a molybdenum source, and completed

the preparation for electroplating.

Conditions of current density and electrolyte temperature used in heat sink processing were reproduced to obtain a gold layer 5 μm thick, and appropriate washing followed by drying yielded samples which were ready for thermocompression bonding.

4.2.4 THE EXPERIMENTAL FACILITY AND BONDING PERFORMANCE

The assembly jig described in Section 4.2.1 was used in conjunction with an Instron model TT-D mechanical press and the complete experimental system is shown in Plate 4.2.4. Although it proved useful to be able to vary and monitor the loading rate during early trials when process limitations were being ascertained, most of the sophistication in the Instron press was superflous, and the equipment chosen mainly for its capacity (maximum load = 10^4 kg), and clean laboratory environment. The asbestos underlayer, and the alignment sphere located between conical recesses in the piston and press cross-head were not included within the original apparatus which suffered from recurrent heater failure associated with excessive heat sinking. The sphere reduced the top area of thermal contact and compensated for any deviation from parallelism between the press faces, while the asbestos provided insulation enabling more accurate measurement of substrate temperature and a reduction in the required heater current.

Derivation of suitable bonding conditions was largely empirical, founded on previous experience of thermocompression

Integral Slice Thermocompression Bonding Apparatus

Plate 4.2.4

ball bonding and conditions which Hambleton had reported to be experimentally satisfactory. Assuming the fusion mechanism to be independent of contact area, it appeared that bonding pressure, being some inverse function of process temperature, required to be increased by a factor of 35 in order to accommodate a reduction in temperature from 300°C to 150°C. While this evidence was in no way definitive, it did indicate a probable operating range of bonding parameters which could be used as a basis for our work.

In order to ascertain any incidence of crystallographic derangement arising from bonding stresses in the samples, high resistivity silicon slices were subjected to normal processing, and finally subjected to a range of thermocompression cycles at a maximum available temperature of 200°C. The general procedure was to assemble a sample and heat sink within the alignment jig and arrange the press cross-head above the load transmission sphere with sufficient clearance to allow free system expansion during the heating phase. An equilibrium temperature of 200°C was achieved in 20 minutes, and the appropriate stress cycle performed.

Microscopic observation revealed only peripheral damage in samples subjected to pressures in the range 2×10^{6} - 1.4×10^{7} $kg.m^{-2}$, and resistivity measurements over sample surfaces by the standard four point probe technique were consistent with pre-process values. Within the specified stress limits, results were unaffected by variation in pressure loading rate from 0.9×10^{4} $kg.m^{-2}s^{-1}$ to 1.8×10^{5} $kg.m^{-2}\ s^{-1}$, and total stress time from 1 min to 20 min. Pressures in

excess of 1.6×10^{7} kg.m^{-2} caused frequent sample disintegration, and this was attributed to uneven deformation of the copper heat sink at the edge of the sample, which had caused only localised fragmentation under less stress.

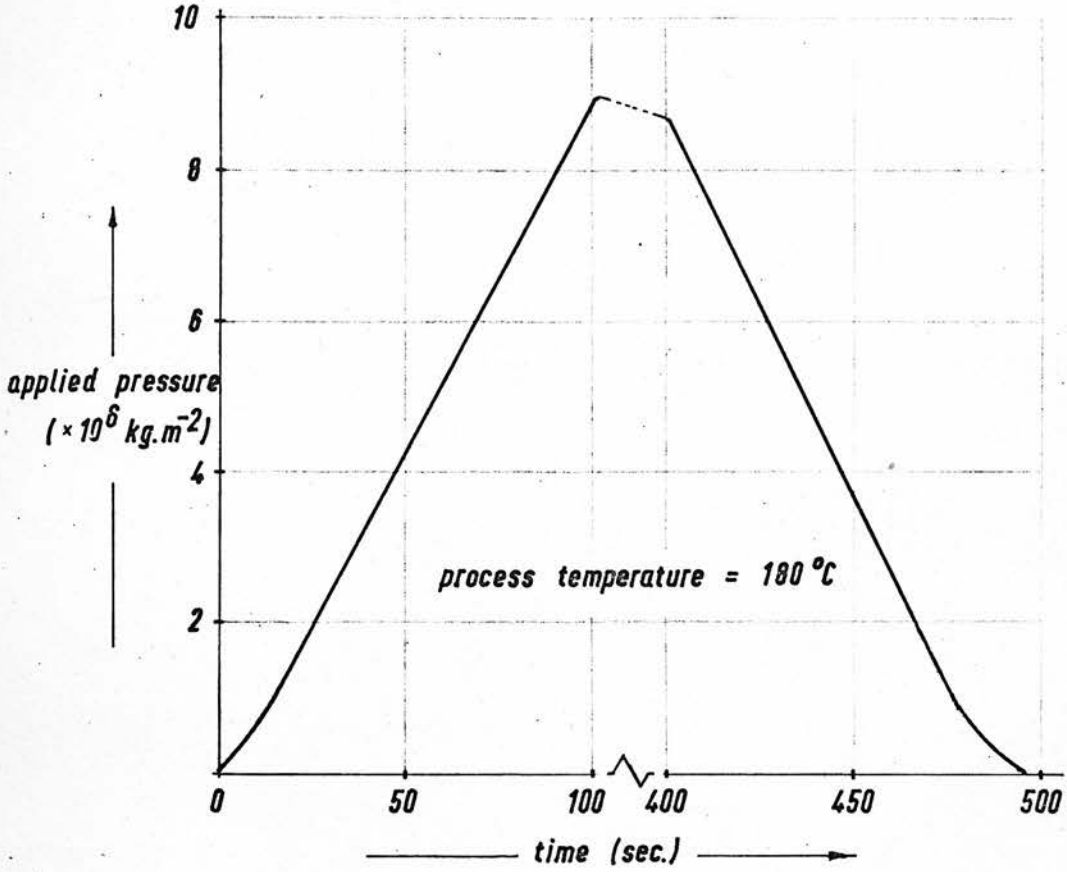
This experiment served to indicate the inevitability of some sample fragmentation resulting from non-uniform peripheral stress, and a maximum pressure which could be applied for minimal damage. Although such a limitation seemed acceptable for our requirement, it was necessary to direct development towards a lower temperature process in order to reduce heater vulnerability.

A standard pressure loading rate of 9×10^{4} kg. m.$^{-2}$.s^{-1}, an operating temperature of 180^{o}C, and a 5 min. stress period at maximum load were incorporated in the thermocompression test schedules where an applied pressure was increased from 2×10^{6} kg.m^{-2} in successive experiments until sample adhesion became evident. The stress cycle shown in Fig. 4.2.4 was found to provide a degree of fusion from which sample detachment could only be achieved by fragmentation at the slice parameter. Furthermore, it was observed that forced separation occurred at the aluminium - nichrome interface, and that the gold plated layer formerly attached to the sample maintained adhesion to the heat sink.

Several samples were processed according to this schedule to confirm reproducibility and provide material for the succeeding operation.

4.3 DEVICE DEFINITION AND ISOLATION

4.3.1 /



Thermocompression Stress Cycle for Slice Bonding

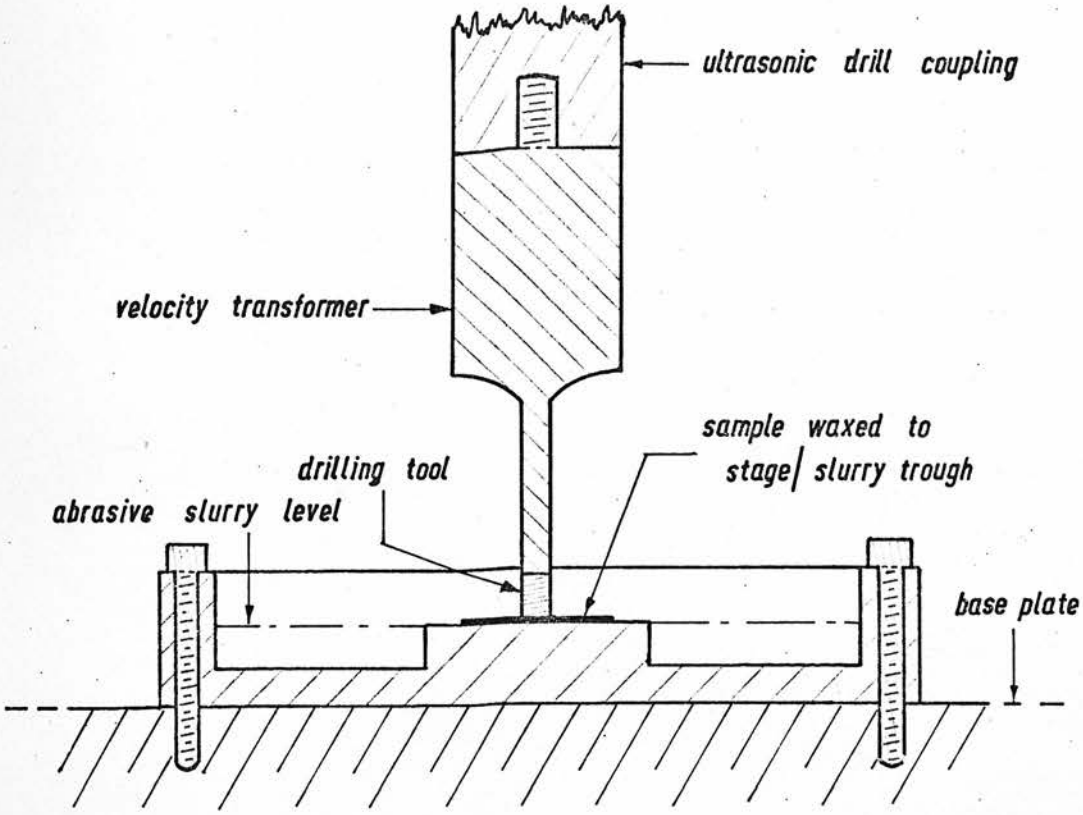
Fig. 4.2.4

4.3.1 AN ULTRASONIC ABRASIVE TECHNIQUE

The potential of ultrasonic abrasive device definition and isolatien in reducing process time and eliminating the need for photolithographic or chemical etching facilities was considered worthy of experimental assessment. In addition to providing information concerning the applicability of this method to complete device separation, such an investigation could make a contribution to semiconductor component manufacture, where a requirement exists for small quantity diode production without large capital investment.

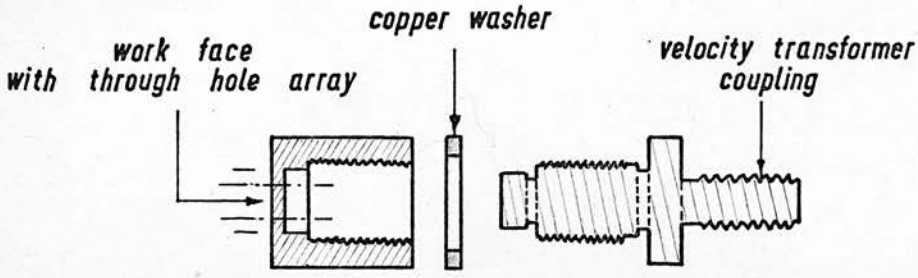
The equipment available for this work was a Kerry ultrasonic drill Type KDS 100 provided with a power unit capable of manually variable output over the frequency range 18-20 kHz. System design problems were concerned with achieving resonance within the available frequency range, ensuring maximum energy transfer to the abrasive slurry, and preparing a suitable tool to optimise drilling rate and minimise tool wear.

The work region schematic of Fig. 4.3.1.1. includes a velocity transformer which matched the equipment energy transmission surface area to that of the tool, a drilling tip, and sample holder with abrasive reservoir. The transformer, made from aluminium bronze in order to minimise losses, was reduced to the appropriate working diameter at the mid-frequency quarter wavelength, and shortened over the final section to enable a resonance condition after tool attachment. Effective energy



Work Region of Ultrasonic Abrasion Apparatus (in section)

Fig. 4.3.1.1



Two Part Construction of Cylindrical Ultrasonic Abrasive Drilling Tool (in section)

Fig. 4.3.1.2

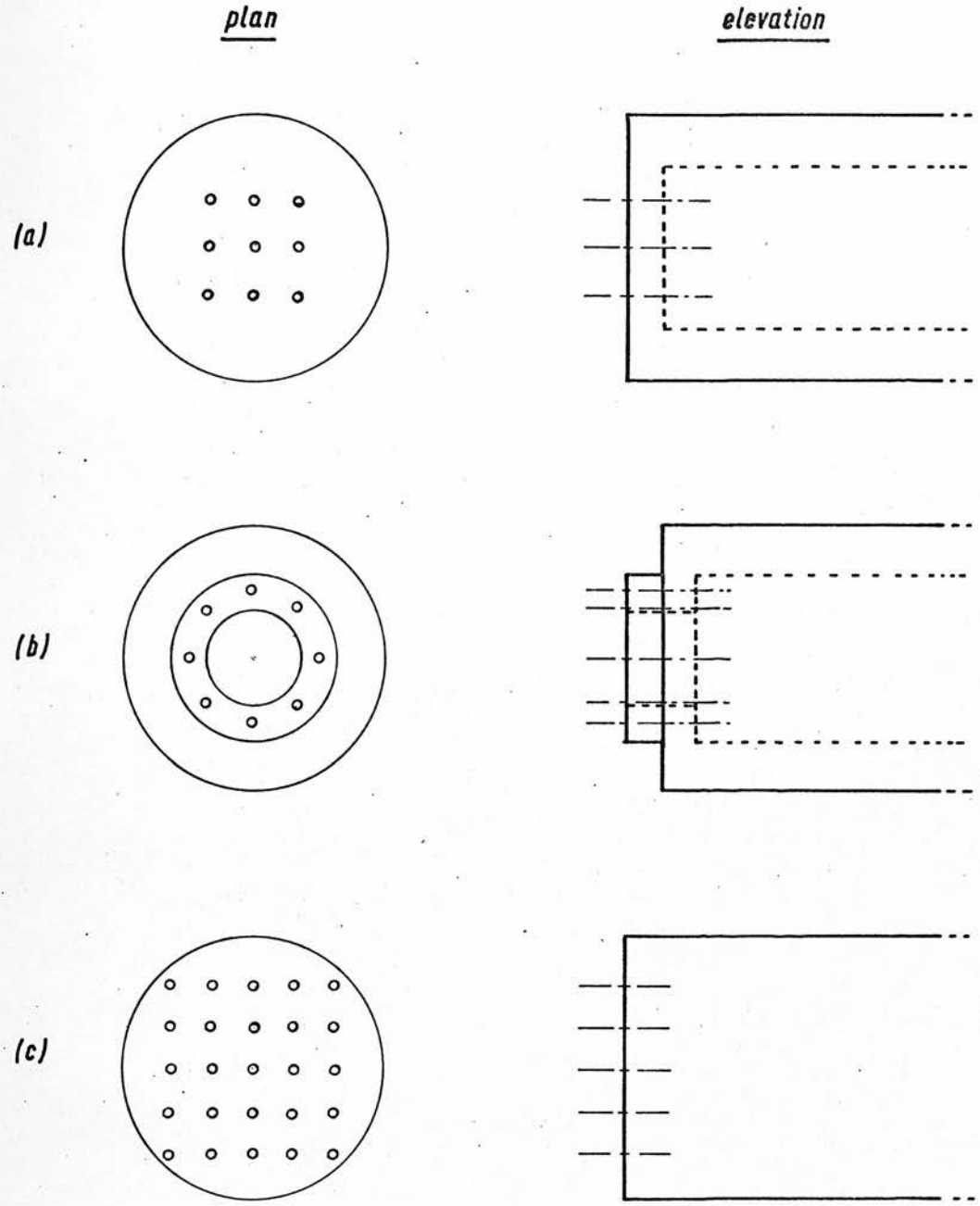
transmission to the abrasive slurry was largely controlled by the intimacy of contact between velocity transformer and tool, however, until the most suitable tool material and configuration had been chosen, it was necessary to be able to separate the components, and the design of Fig. 4.3.1.2 was developed for initial experiments. The two part tool construction permitted close inspection and cleaning after each drilling operation, a facility which proved invaluable during the trial period where tool wear was prolific and the incidence of blockage high.

An additional design consideration was the difference in sound velocity between the various tool materials, which necessitated independent calculation of each tool length in order to ensure resonance.

First attempts to etch cylindrical structures in silicon were to be made using a mild steel tool drilled with a 3 x 3 matrix of 125 μm diameter holes as depicted in Fig. 4.3.1.3a. The hole spacing was chosen to be 0.16 cm between centres so that sufficient heat sink area would subsequently be available for device separation.

Some difficulty was experienced in realising the required tool profile since normal workshop activities did not include such close tolerance and delicate work: accordingly, a laboratory process had to be developed to meet our requirements.

Surfaces of a tool 'blank' were ground to provide



Ultrasonic Drilling Tool Configurations

(scale = 4 times full size)

Fig. 4.3.1.3

assembly references, and extreme care exercised in securing the component to an x-y platform positioned on the work table of a high speed precision drill. This lengthy operation was essential to ensure that the direction of hole recession from the work face would coincide with ultrasonic drilling motion. Machining was performed with a watchmaker's drill constantly lubricated with paraffin, and manually loaded from a micrometer coupling while the work region was observed under a microscope.

Despite numerous attempts with different loads, energy levels and abrasive materials, it was not possible to ultrasonically drill through 0.3 mm thickness of silicon and retain cylindrical structures within the tool recesses. Machining times in the region of 5 min. were excessive and contributory to specimen fracture which commenced after 0.1 mm tool penetration.

In an attempt to reduce the amount of silicon being abraded, the tool profile of Fig. 4.3.1.3b was adopted. This modification had the desired effect, and a machining time of 30 s. obtained when a water and 400 grit silicon carbide abrasive slurry was used in conjunction with a tool load sufficient only to maintain penetration. It was particularly important to arrest the process immediately the drill arrived at the silicon - gold interface, since prolonged exposure of formed components to ultrasonic energy in the region of heat sink attachment caused separation and disintegration. Providing this precaution was taken however, a yield of 90% could be achieved quite readily.

Close examination of the tool work face after one drilling operation revealed considerable erosion around the recess perimeters. This alteration in profile had the undesirable effect of modifying the silicon shape in a complementary way; and consequently an alternative tool material had to be sought. Degradation could not be eliminated although KE 672 tool steel, proving superior to stainless or silver steels, gave a ten times increase in durability over the original material.

Once the decision to use KE 672 steel had been taken, a revision of equipment construction was implemented in an attempt to drill a 5 x 5 component matrix. After oversize machining, a tool blank and appropriate velocity transformer were butt joined at 630°C using a Eutecrod 1801 alloy developed by the Castolin Eutectic Company to enable low temperature bonding by diffusion of base materials into a proprietary filler. The resultant assembly was ground to final size, and the tool face lapped using a 3 μm diamond compound. Recess drilling was performed as described earlier, and final lapping used to ensure a true work face. The matrix configuration is shown in Fig. 4.3.1.3c.

This arrangement allowed more rapid drilling than the earliest tool design despite an increase in total silicon removal. Nevertheless, the 3 min. operation caused significant sample disintegration and the anticipated array could not be realised. Furthermore, a scanning electron micrograph of a typical surviving structure, reproduced in Plate 4.3.1.1 showed two undesirable

Ultrasonically Drilled Diode

(x180)

Plate 4.3.1.1

(a)

Surface Finish

of Above Diode

(x1400)

(b)

Surface Finish

of Above Diode

after Chemical Polishing

(x1400)

Plate 4.3.1.2

Drilled / Etched Array

(x40)

Single Diode

(x200)

Plate 4.3.1.3

process features. The curved profile which existed in the vicinity of the heat sink was a product of tool erosion discussed earlier, and the irregular silicon surface in the junction region would have caused localised low voltage breakdown for simple geometric reasons, and excessive leakage from gross contamination. The general surface appearance of the silicon, detailed in Plate 4.3.1.2a, was also too granular, being a function of the grit size used during erosion.

In an attempt to rectify these problems, finer grades of silicon carbide, alumina powder, and diamond pastes were substituted for the original 400 grit, and drilling experiments performed using water, vegetable oil and mineral oil media. In each case, process time became excessive with consequent sample disintegration.

Since ultrasonic techniques appeared ineffective in producing the requisite silicon surface, a chemical polishing process had to be introduced. The structure shown in Plate 4.3.1.1 was submerged in a room temperature solution of

4 parts acetic acid
+ 6 parts nitric acid
+ 3 parts hydrofluoric acid
by volume of analar chemicals,

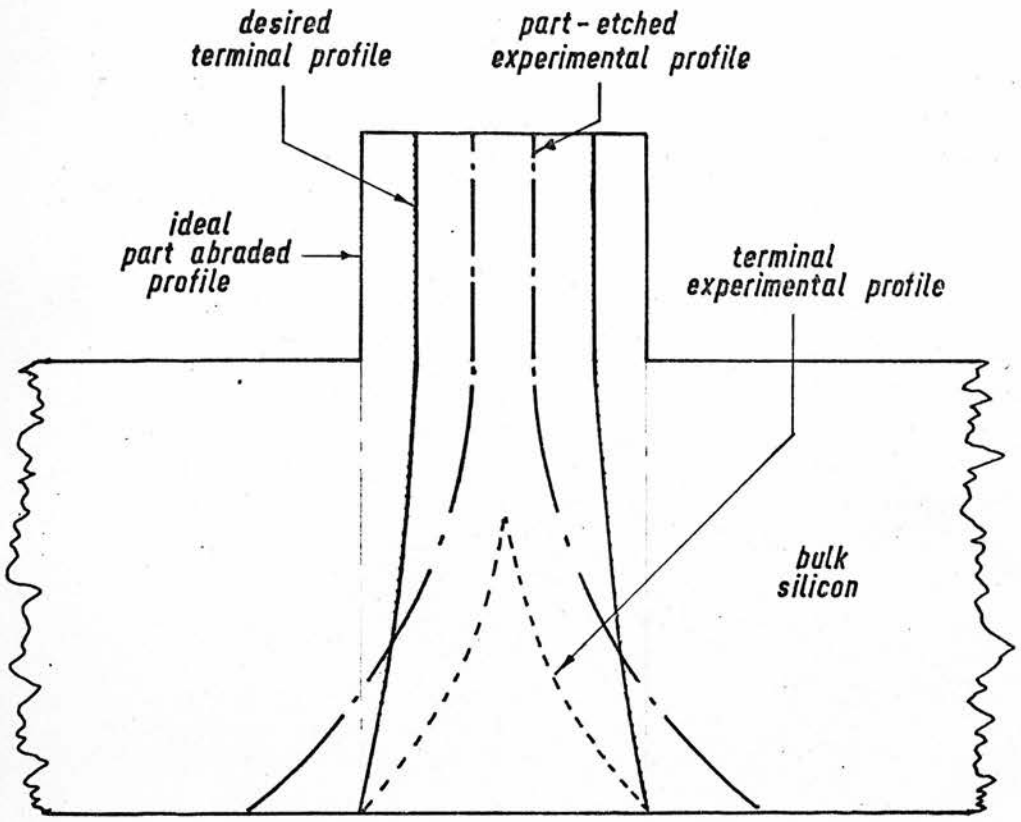
and the resultant almost featureless contour illustrated in Plate 4.3.1.2b for comparison. The effectiveness of this treatment was obvious, and thought directed towards the problem of low yield

associated with matrix drilling.

Once the need for chemical etching had been acknowledged, it seemed reasonable to exploit any advantages which the process might have.

It will be recalled that lateral etching in a chemical process was the inhibiting feature in its use for device isolation from a 0.3 mm thick slice. Nevertheless, used in conjunction with ultrasonic abrasion capable of preliminary device definition to a depth of 0.1 mm as already demonstrated, chemical etching could be used to remove the remaining unwanted substrate material as indicated in Fig. 4.3.1.4. Clearly the success of this combination was dependent on reaction behaviour since devices were completely exposed to etchant throughout the process.

During preliminary work carried out to determine the most suitable etching cycle, it was found that slow etches with a high nitric acid concentration, used primarily to give a uniform silicon surface, caused an unacceptable degree of lateral etching. Since a chemical polishing technique had already been satisfactorily evaluated, emphasis was placed on obtaining a solution with low lateral etching properties and a reaction rate which would not obscure observation of progress through excessive gaseous evolution. This compromise was necessary to ensure that etching could be arrested as soon as the required component profile had been achieved.



Combined Ultrasonically Abraded / Chemically Etched Mesa Profile

Fig. 4.3.1.4

The best results which could be obtained practically using a solution of

3 parts acetic acid
+4 parts nitric acid
+3 parts hydrofluoric acid
by volume of analar chemicals

are shown in Plate 4.3.1.3. It will be noticed that preferential vertical etching had not been as significant as anticipated, and although the process was stopped prematurely, the silicon profile had become unsuitable for further development. Such an inherently fragile structure would have given considerable problems during the attachment of a top electrical contact, the thermal and electrical resistance of the narrow portion would have been excessive, and any attempt to continue reduction of the component diameter to 100 μm near the base would have resulted in pinnacle formation.

Although some thought was given to methods of removing the upper portion of such structures, it became obvious that endeavours to incorporate ultrasonic drilling were defeating the original object of providing a simple and inexpensive method of diode formation. Problems had arisen in drilling a small matrix, and it was unlikely that further development would enable complete slice processing in a single abrasive operation. The quality of silicon surface had been restricted by the size of grit required, and chemical polishing became an essential part of the process.

This admission detracted seriously from the advantage of ultrasonic abrasion as a means of device definition and isolation: consequently the formidable problems encountered, coupled with an urgent need to fabricate a functional device, led to the rejection of ultrasonic drilling in favour of a closer examination of chemical alternatives.

4.3.2 A CHEMICAL ETCHING TECHNIQUE

The primary difficulty associated with chemical isolation was the introduction of silicon slice thinning to the fabrication schedule. While, in the interests of minimised sample vulnerability, it was desirable to perform this act after thermocompression bonding, the time taken to develop a suitable thinning technique would have been extended by the need to prepare a considerable number of heat sinks and by the elaborate processing necessary before each attempt at silicon thickness reduction.

As a compromise, it was decided to incorporate slice thinning immediately after gold electroplating when most of the sample processing was complete. This enabled early simulated appraisal to be performed on slices prepared by simple gold deposition without the need for alloying.

In order to obtain suitable diode profiles by chemical isolation, it was necessary to reduce samples to less than 0.1 mm in thickness, and initial experiments were aimed at achieving this by mechanical lapping. Use of a 3μm diamond compound on a

cast iron lap enabled stock removal to a thickness of 0.15 mm before peripheral sample fracture took place. Modifications to the process including reduction in loading, substitution of a Caprate Optical Grade C20 wax for the conventional Apiezon (black) wax, and use of a Microcloth lapping base enabled sample thicknesses of 0.1 mm, but the incidence of success was insufficient.

An alternative scheme which did meet the requirements however, involved preliminary sample lapping to 0.15 mm and subsequent chemical 'float' etching until the requisite stock material had been dissolved. The latter process consisted of placing the sample on a filter paper impregnated with

3 parts acetic acid
+5 parts nitric acid
+3 parts hydrofluoric acid
by volume of analar chemicals

to achieve an etch rate of 50 $\mu m.\ min^{-1}$ and final sample thickness of 40 μm . Suitable protection for the sample active region was provided by the gold overlayer, and a final wash in double distilled, de-ionised water followed by drying in a nitrogen atmosphere prepared the silicon for thermocompression bonding.

It is sufficient to comment that no difficulty was experienced in bonding thinner samples and device fabrication could proceed to chemical isolation.

While devices could have been defined on the silicon

surface using a Kodak K.T.F.R. photoresist capable of withstanding a limited period of etching, this technique was rejected on consideration of the later problem of achieving ohmic contact to the n^+ region. It was anticipated that vacuum deposition of gold would be used to accomplish this, and that devices resulting from a chemical etching cycle would possess mesa surfaces entirely exposed to the metal vapour stream, giving rise ultimately to a short circuit condition.

A possible solution to this problem was to deposit the gold layer in advance of device definition and reproduce the necessary diode pattern in this overlayer by a photoresist technique. In practice, gold was evaporated from a molybdenum source within the vacuum system to establish a 400 Å thin film on the n^+ sample surface which was maintained at 150°C during deposition. The chamber pressure was $< 5 \times 10^{-5}$ torr and film thickness, although uncritical, was estimated from a calibrated Speedivac quartz crystal monitor. On the assumption that gold areas defined during a photolithographic stage would provide a satisfactory mask during mesa etching, it was permissible to use AZ 1350 photoresist during pattern delineation and residual gold could be removed in a solution of

300 g. potassium iodide dissolved in 1 ℓ. de-ionised water
+ 100 g. iodine
+ additional water to a total volume of 2ℓ.

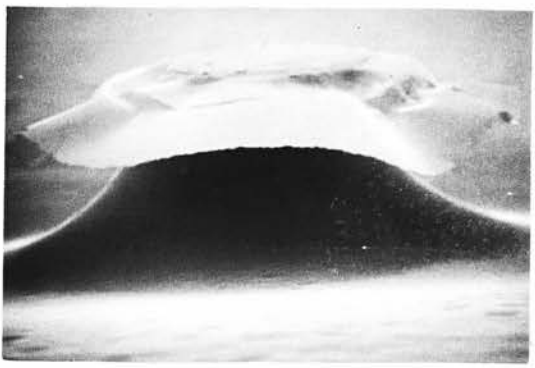
Unfortunately a complication arose in the application of

this technique, as shown in Plate 4.3.2. During chemical isolation, lateral etching gave rise to a fragile, unsupported film of gold beyond the upper surface of the mesa, and it was conceivable that such vulnerable overhanging material could introduce a fault condition in small devices: accordingly, remedial processing had to be found.

The most suitable modification to the fabrication schedule involved two stage photolithography where a matrix of gold dots established during the first phase was covered by a concentric K.T.F.R. pattern suitably increased in area to compensate for under-etch. This corrective measure was impractical however, since the available mask alignment facility was incapable of the precision required to obtain correct registration of the second pattern.

The realisation that further laboratory time would have to be spent on an aspect of diode fabrication which had already consumed a disproportionate part of the project duration gave cause for concern. Clearly, an immediate solution had to be found which would enable completion of the fabrication schedule and the release of some devices for dynamic testing.

A close examination of thermocompression ball bonding revealed a consistency in bonding area which, by use of 0.05mm thick gold wire and a load of 20g. at 250^{o}C, could be arranged to be 0.125 mm in diameter. This process offered the opportunity of forming mesa diodes where the ball bond provided adequate



Chemically Etched Mesa

(Gold Thin Film Masking)

(x300)

Plate 4.3.2

Copper Device Mount

(x14)

Plate 4.4

masking during an etching operation, and the attached wires facilitated electrical contact during device test.

Following an integral slice mounting stage therefore, the n^+ silicon sample surface was coated with gold as described earlier in this section, a regular matrix of ball bonded wires attached, and the residual gold layer dissolved. Mesa etching was performed in a solution of

1 part acetic acid
+2 parts nitric acid
+1 part hydrofluoric acid
by volume of analar chemicals

to yield devices approximately 100 μm in diameter which were washed in boiling water (double distilled, de-ionised), dried in nitrogen and passed to the separation phase.

4.4 DEVICE SEPARATION

During the study of ultrasonic abrasion, some time was devoted to an appraisal of copper machining. The malleability of this metal proved a severe impediment however, and particle removal was secondary to a displacement process whereby residual material accumulated within the tool recesses and around the work area perimeter. It was considered unlikely that a prepared diode would be able to withstand such distortion of the heat sink surface, or indeed exposure to ultrasonic energy throughout a protracted machining period.

Since ultrasonic abrasive equipment was no longer required in earlier processing, it was unprofitable to persevere with such a basically unsatisfactory technique, and effort was directed to more conventional machining methods.

Experiments were conducted on sectioning a copper sample using a slitting saw mounted in a horizontal milling machine. Specimens were waxed to an intermediate mounting base which was attached to the milling table, and various cutting actions investigated.

In the interests of minimising waste which would reduce the diode yield per slice, saw blades were limited to 0.1 mm thickness. Excessive tool wear proved the most restricting influence on machining technique, and sawtooth fracture was a common occurrence during early work. It was eventually established however, that 'climb' milling with a 5.7 cm. diameter blade (4 teeth per cm.) at a tool speed of 920 r.p.m. and sample feed of 1.1 $cm.min^{-1}$ would permit cuts of 1.50 mm in depth, providing the cutter was continuously lubricated with AKA wax.

By superimposing two such cuts, it was possible to machine a groove to within 0.10mm of the heat sink reverse surface, and repetition of this processes at appropriate intervals resulted in an array of orthogonal channels creating a matrix of rectangular pillars 1.50mm square in section, 3mm deep, and held together by a continuous base copper layer of 0.10mm thickness.

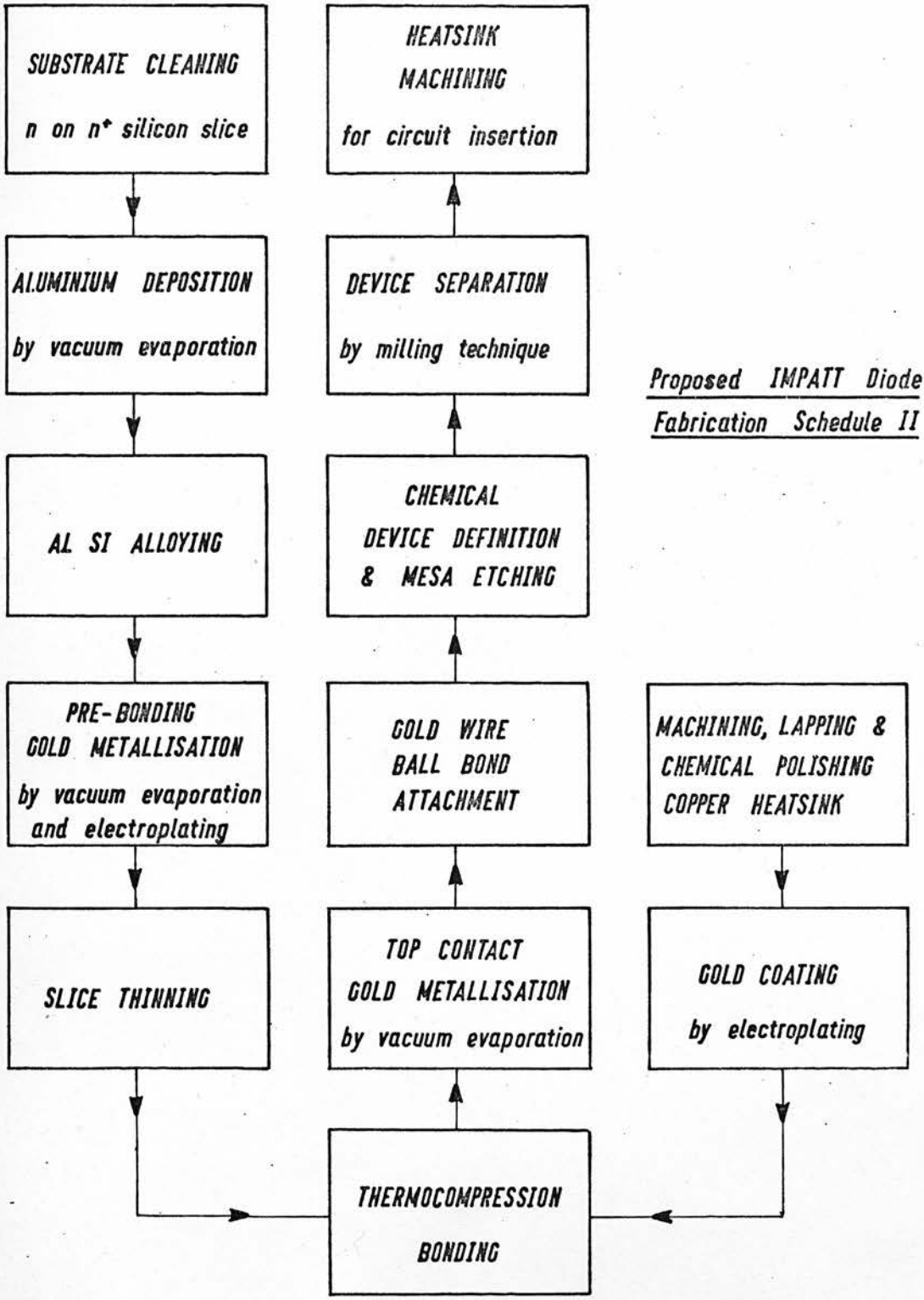
On removal of the sample from the intermediate platform, heat

sink separation was performed manually, and individual segments machined and threaded to yield a component, illustrated in Plate 4.4, which could be inserted directly into the test circuit.

It remained therefore, to repeat this operation with a sample which had undergone the entire process schedule, and devices would be available for dynamic testing. The problem of diode protection during this step in manufacture was overcome by coating components with a layer of Lakeside resin which was impermeable to cutting lubricant and offered a degree of physical resistance to indelicate treatment. Removal of the resin was achieved by individual sample immersion in ethyl-methyl-ketone, and device fabrication terminated by rinsing components in iso-propyl alcohol and allowing a drying period in a nitrogen atmosphere.

4.5 A COMPLETE DEVICE FABRICATION SCHEDULE

This chapter has described the progression of device development through every aspect of processing. Such a catalogue of alternative processes, and indeed changes in constructional order, tends to conceal the configuration ultimately chosen. By way of a summary therefore, Fig. 4.5 presents the final fabrication schedule used to realise p^+nn^+ diodes which could be dynamically tested for IMPATT properties, i.e. for microwave generation.



Proposed IMPATT Diode
Fabrication Schedule II

Fig. 4.5

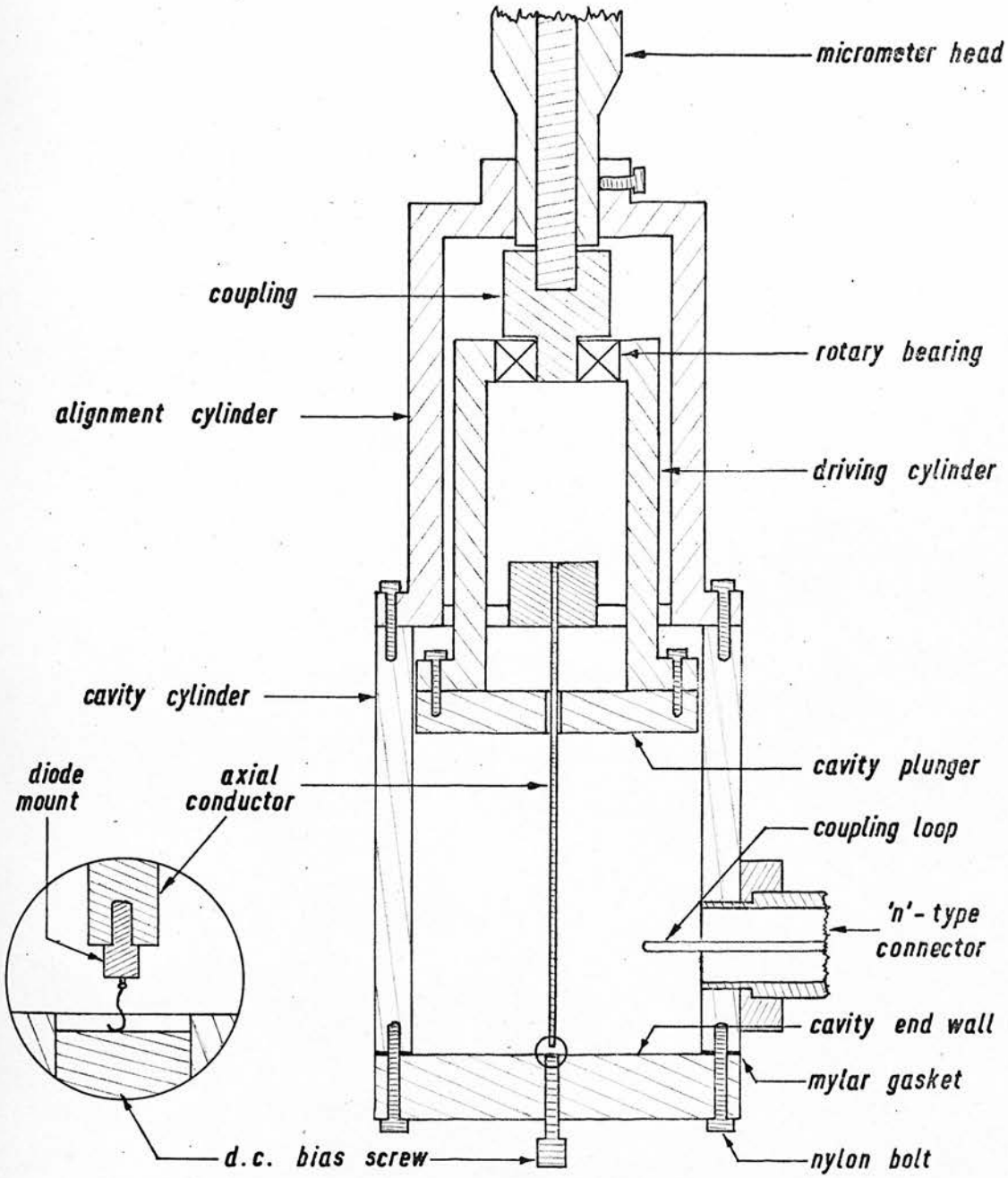
CHAPTER 5

A TEST FACILITY FOR MICROWAVE DETECTION, DYNAMIC DEVICE TESTING, AND CONCLUSIONS.

5.1 DESIGN AND CONSTRUCTION OF A RIGHT CIRCULAR CYLINDRICAL CAVITY RESONATOR.

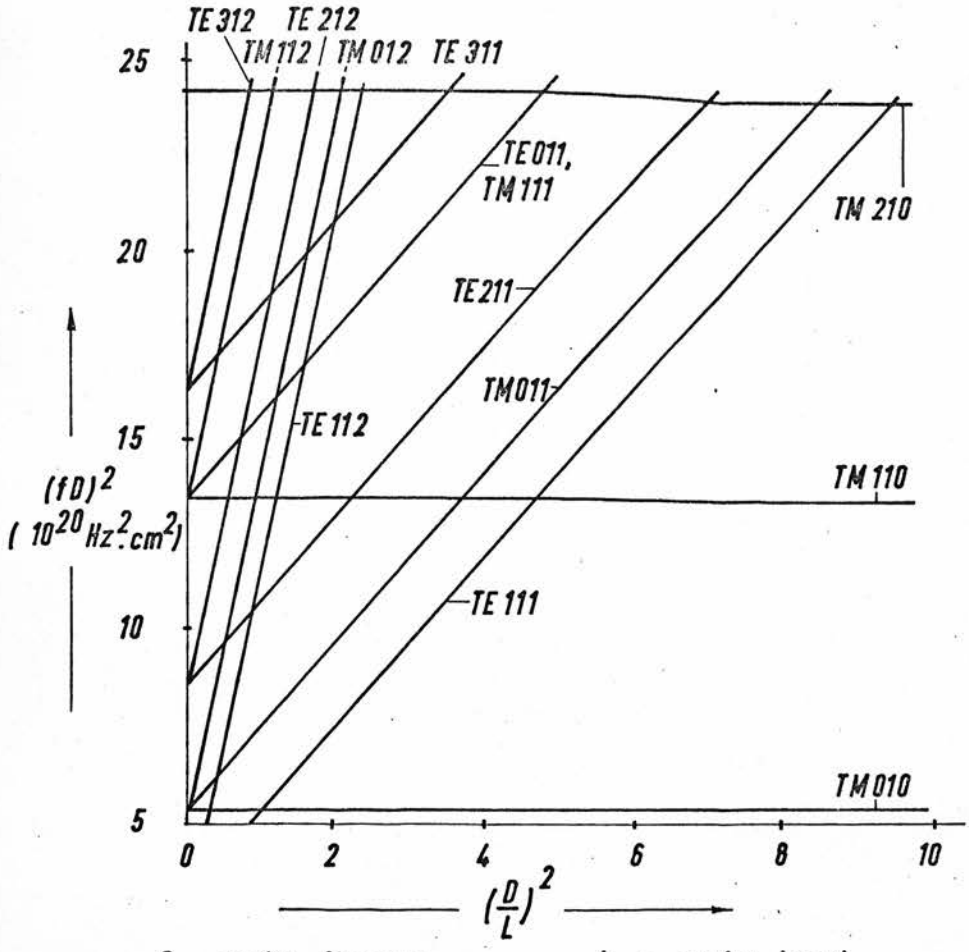
The design aim of a versatile IMPATT diode oscillator was to conceive a low loss, variable frequency and singly resonant circuit which could be calibrated to give unambiguous frequency measurement over the range imposed by limitations in device processing. Such an ideal requirement was approximated by the use of a right circular cylindrical cavity resonator designed to operate at 9.5 GHz in the TE 011 mode, with a movable end wall advanced by a conventional micrometer head to give $\pm$ 0.5 GHz adjustment about this centre frequency.

Fig. 5.1.1 is a schematic of the configuration constructed for this inherently high Q excitation mode, where end wall currents were circumferential and hence compatible with the electrical isolation required to accomodate a movable plunger and an isolated end wall supplying d.c. bias to the generating diode. A desirable feature of this discontinuous structure was the elimination of radial end wall currents, and thereby, the suppression of a TM111 mode which would otherwise have reduced the effective cavity Q by energy transfer . This situation may be simply understood by reference to Fig. 5.1.2 a,b (extracted from



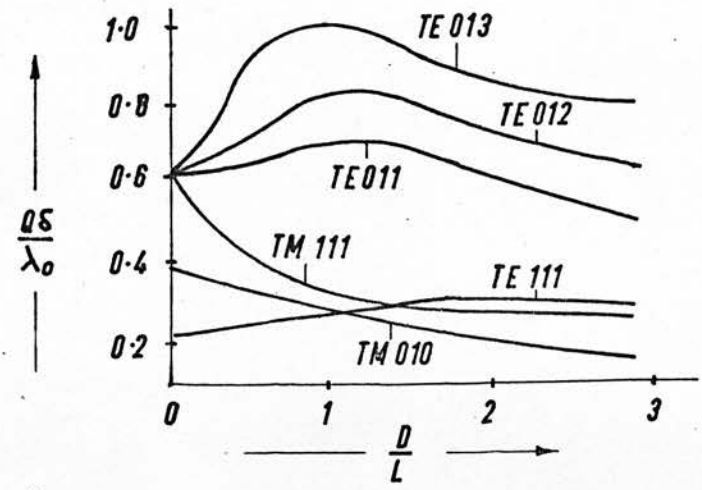
Typical Axial Section of TE 011 Cavity Assembly

Fig. 5.1.1



D = cavity diameter L = cavity length

(a) Mode Chart



δ = skin depth

(b) Graph of $\frac{Q\delta}{\lambda_0}$ against $\frac{D}{L}$.

Fig. 5.1.2

a text by Montgomery[49] from which it will be noticed that TM 111 excitation has the unique property of being resonant at the same frequency as TE011 for a given cavity geometry, and is of intrinsically lower Q.

A further point arises from these graphs, concerning the suitability of higher TE0ln modes. While these have still higher Q factors, they require larger cavity volumes, and it was unlikely that machining accuracy could be maintained to enable the theoretical benefit of larger construction.

The central conductor, tapped to receive the diode assembly and providing continuity for the d.c. supply, was constructed from thin steel rod and physically isolated from the cavity walls. This minimised interference with the desired field pattern, and the lossy nature of steel at microwave frequencies assisted in suppression of a TEM mode.

The method of cavity coupling was largely dictated by electrical and mechanical constraints. Because of obvious problems associated with attachment to the d.c. biased and wall or the movable plunger, it was necessary to locate a coupling fixture on the cylindrical cavity surface. Circulating currents caused to flow in a coupled system give rise to dissipative losses which lower the cavity Q: accordingly, a site was chosen from field considerations and the desire to use a loop mechanism, which unlike a probe, is capable of providing useful output without the need for close coupling.

In the basic construction of a TE 011 cavity, it was possible to adopt a treatment devised by Morton[50] during previous work in this Department, and modify his technique to accomodate the coupling and biasing aspects peculiar to our program.

The peak Q-factor for a TE 011 operational mode in a right circular cylindrical cavity occurs when the diameter: length ratio is unity; and the cavity resonates at 9.5 GHz when its diameter is 4.25 cm. Since final cylinder machining was to take the form of a ball burnishing process, where a highly polished steel sphere was forced through the undersize component, it was necessary to adopt a cavity diameter of 4.45 cm. (1.75 in) being the nearest available sphere size and maintaining optimum cavity performance within the 9-10 GHz range.

The cylinder was machined from hard drawn, high conductivity (H.D. H.C.) copper to an inside diameter of 4.3 cm. and a wall thickness of 0.635 cm., while care was taken to minimise internal surface irregularities. Ball burnishing was accomplished under a hydraulic press once the steel ball and copper cylinder had been lubricated with silicone oil. The cavity end plate and plunger were machined from similar H.D. H.C. copper, the plunger diameter being reduced to provide 0.63mm. clearance around the cylinder interior, and drilled to allow similar clearance for the central conductor. Both plates were subsequently lapped to a surface roughness $<1500A^{\circ}$ measured on a Talysurf, and together with the cavity cylinder were annealed at

600°C for 1 hour in an argon atmosphere as an attempt to obtain uniform surface impedances of all cavity walls. This procedure encouraged the growth of copper crystallites which roughened the surfaces and necessitated the application of a mild abrasive to reinstate an irregularity level below 1500Å .

The position of an aperture in the cylinder for loop insertion was arranged to permit maximum variation in coupling with the axial magnetic field at 9.5 GHz. This was ascertained from the equation for axial field strength of the TE 011 mode in a right circular cylindrical cavity, which can be reduced to the form

$$H_z = 2jH_1 J_0 (3.832 \frac{r}{a}) \sin \frac{\pi z}{d} \qquad 5.1.1$$

where H_z = axial magnetic field strength, r = radius variable, a = cylinder radius, z = axial distance along the cylinder such that cavity end walls exist where $z = 0$ and $z = d$.

Because of the sinusoidal variation of H_z with respect to z, maximum axial field strength will occur at $z = \frac{d}{2}$, i.e. midway along the cavity. At 9.5 GHz, this corresponded to 1.6 cm from the end walls.

The aperture was machined to accept a suitably tailored female n type co-axial connector fitted with the appropriate coupling loop. Since the function $J_0(r)$ changes sign at $r = 2.45$, it will be seen that loop penetration had to be

restricted to the region $a < r < \frac{2{\cdot}45}{3{\cdot}832} a$ in order to avoid coupling cancellation through the change of field direction: in practice, a penetration of 1 cm and a circular loop area of 0.79 cm^2 were used. Connector mounting was arranged to permit loop rotation about the transverse axis, thereby enabling simple reduction of coupling and hence cavity loading if required.

Other constructional details were concerned with cavity assembly. The mechanism of plunger advancement from the micrometer drive involved two concentric aluminium alloy cylinders: the outer, being directly attached to the cavity cylinder, formed a mount for the micrometer head; while the inner coupled the plunger and drive via a ball race to enable linear motion. The interposition of a cruciform and boss across the end of the cavity cylinder provided a guide for the plunger motion and a fixture for the axial conductor. Electrical isolation of the end plate was arranged by inserting a 25 μm thick mylar gasket between plate and cylinder which were subsequently secured using nylon bolts, and electrical contact with the diode achieved by the introduction of a screwed copper rod along the cavity axis.

5.2 CAVITY CALIBRATION

While the resonant frequency of the cavity could be predicted theoretically from geometric considerations, and a table of frequency against micrometer postion generated, it was not practical to assess the effect of coupling, an axial conductor, or inaccuracies in machining. Accordingly, the calibration

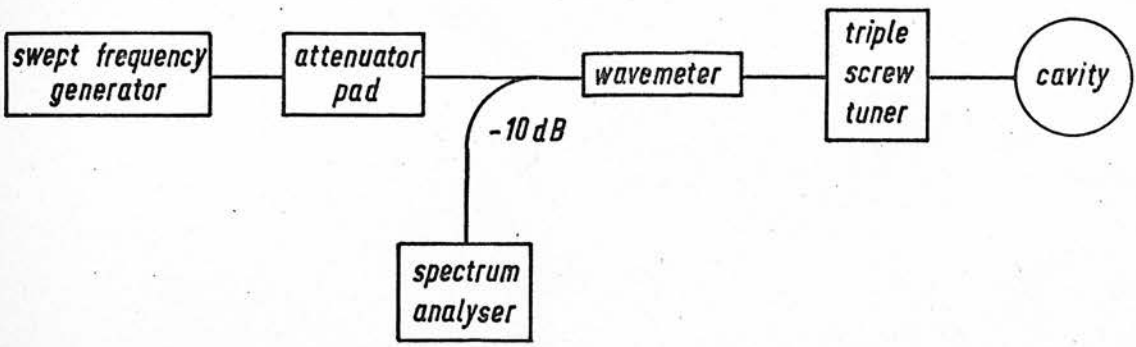
circuit of Fig. 5.2.1 was established to ascertain the presence of resonant modes within the cavity dynamic range.

A swept X-band source was applied to the cavity, and reflected signals detected on a spectrum analyser. (Because of the ambiguous nature of non-real time analysis, it was necessary to incorporate a wavemeter in the circuit for confirmation of operating frequency). Such observation enabled adjustment of system matching, and subsequent resonance calibration was achieved from screen traces similar to that shown in Plate. 5.2

Several of the modes indicated in Fig. 5.1.2.a to be within the dynamic range of the cavity were noted, however only the TE212 mode appeared comparable in amplitude with the TE 011. Both excitations are plotted graphically in Fig. 5.2.2 as frequency functions of cavity length, and the cavity Q for TE 011 operation is indicated in Fig. 5.2.3.

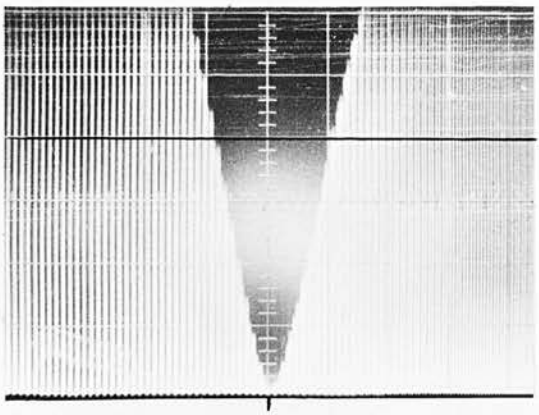
5.3 THE D.C. BIAS CIRCUITRY

Since diodes were required to operate in an avalanche mode,it was necessary by way of device protection to impose current limitation on the d.c. source. Furthermore, it has been shown in equations 2.1.18 and 2.1.27 that diode dynamic impedance is a function of current density, and demonstrated by the fact that a degree of oscillator tuning can be achieved by adjustment of the bias level: while such adjustment was considered desirable in a power supply design, care had to be taken to maintain a high



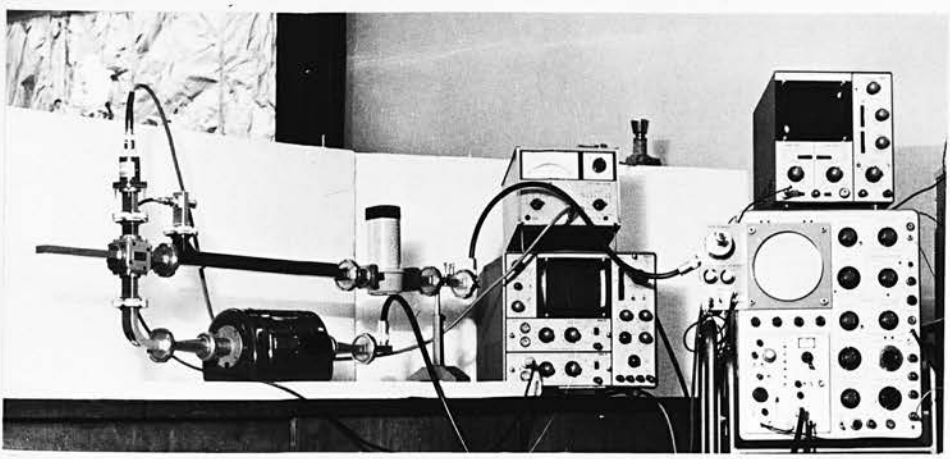
Cavity Calibration Circuit

Fig. 5.2.1



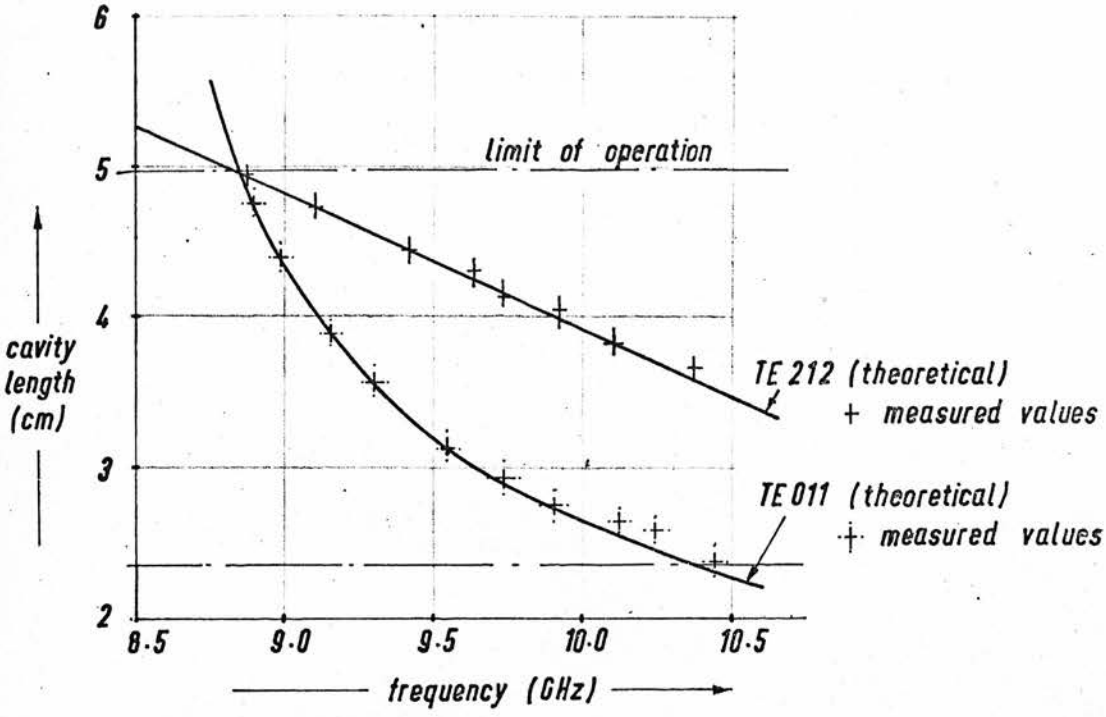
Spectrum Analysis for Q - Measurement

Plate 5.2



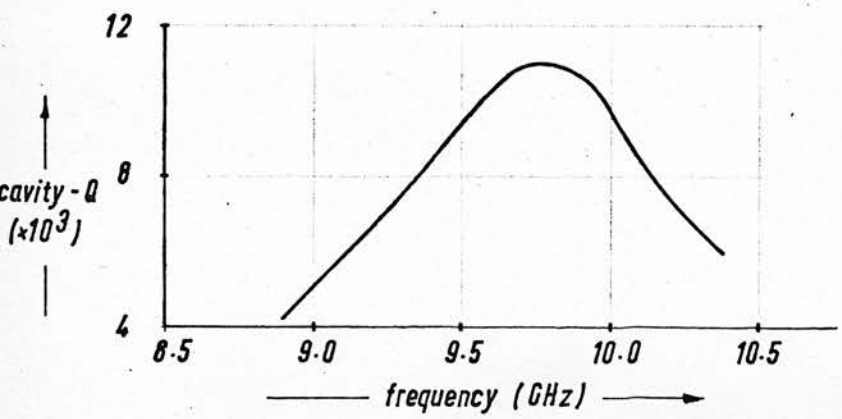
Microwave Test Equipment

Plate 5.4



Cavity Geometry for TE 011 and TE 212 Modes

Fig. 5.2.2



Graph of Cavity - Q against Frequency for TE 011 Mode

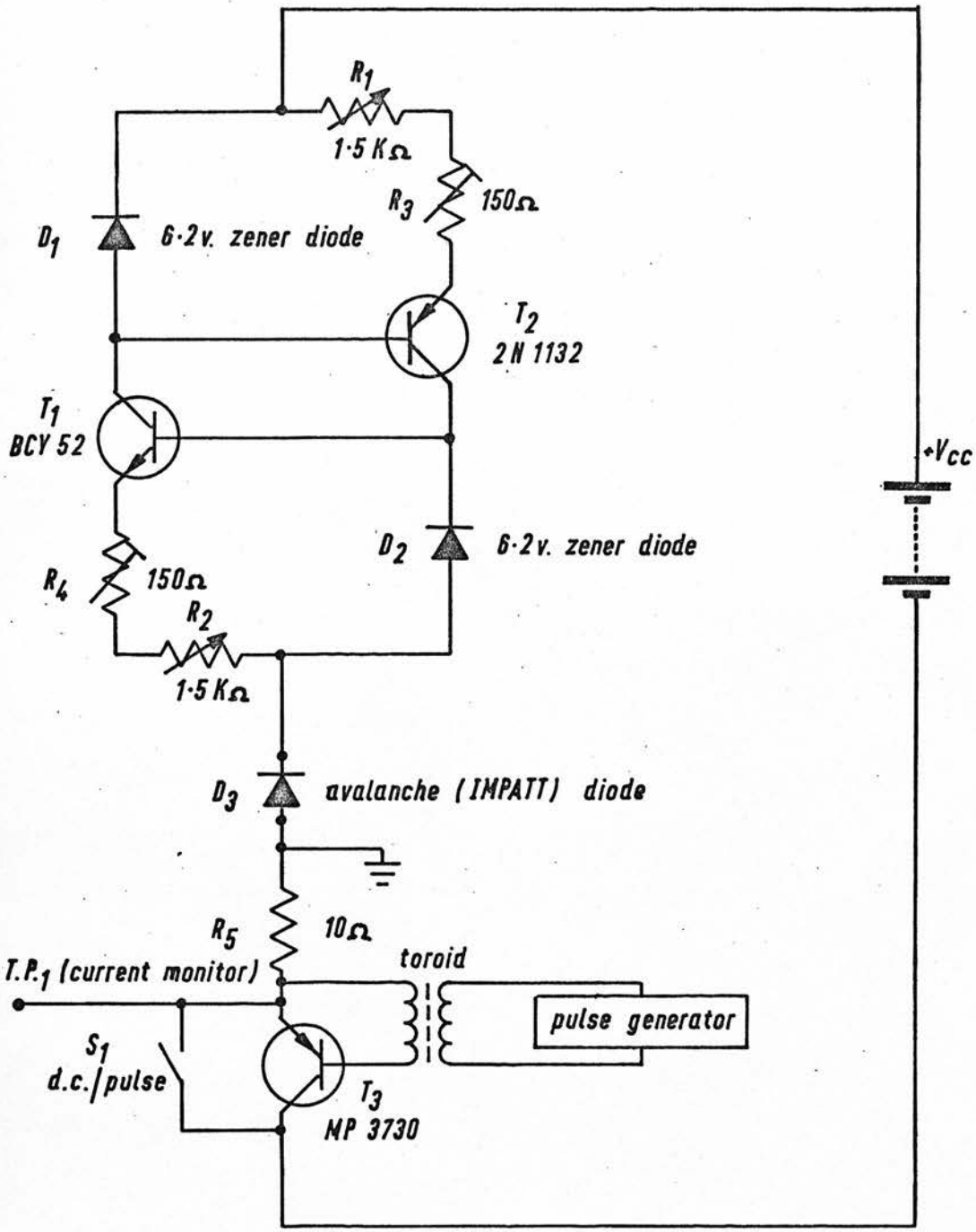
Fig. 5.2.3

degree of current regulation for narrow band operation. Finally, provision had to be made for a variable duty cycle in order that the mean loading could be increased gradually from a safe operating level towards the device limit.

These requirements were met by the circuit design shown in Fig. 5.3 to consist of a coupled constant current supply feeding a transistor switch, the avalanche diode, and a monitoring resistor. The most important feature of controller design lay in the supply to each reference diode, which was a constant current source enabling a total output impedance of approximately 100 MΩ over the operating range 7-100 mA. with applied voltages between 15 and 50v.

Initial setting up was performed with R_1 and R_2 in the minimum position while presets R_3 and R_4 were adjusted to provide an upper limit of 50 mA. from each controlling section. Subsequent variation of R_1 and R_2 enabled normal operation over the limiting range 7-100 mA.

A duty cycle was determined by the switching action of transistor T_3 under the drive of a square wave voltage pulse of the desired duration and p.r.f., coupled from a commercial generator to the base-emitter circuit via isolated windings on a toroid. This apparently cumbersome technique was necessitated by the imposed electrical earth location, and a desire to monitor diode current on a general purpose oscilloscope already equipped with an earth terminal. In the choice of a suitable switching transistor, it



D.C. Bias Circuit

Fig. 5.3

was essential to ensure a specification including $V_{CB\ max} > V_{CC}$ so that the device would withstand an open circuit condition.

Current monitoring was achieved by measuring the voltage developed across a 10 ohm resistor which gave the equivalence 0.1v ≡ 10 mA. with a linear correlation.

5.4 A TEST CONFIGURATION FOR MICROWAVE OBSERVATION

Since confidence had been established in the operation of both the cavity and diode bias supply unit as individual components, it was possible to assemble a test facility capable of microwave operation.

The complete equipment, shown in Plate 5.4, contained the coaxial cavity, output coupled to X-band waveguide in a manner identical to that used for calibration experiments, where a screw tuner optimised system matching. An X-band wavemeter was inserted adjacent to the waveguide input, enabling unique determination of propagation frequency; while the generation level was constantly monitored on a detector diode at the end of a following -10 dB coupler. The main waveguide transmitted incident output via a switch to either a thermistor load forming the input to a mean level power meter, or to a spectrum analyser through a protective calibrated variable attenuator.

This combination allowed measurement of mean output power, instantaneous X-band output and the power frequency distribution. In general, spectrum analysis was confined to the region

of maximum signal, while power summation was performed by means of the detector diode and corroborated by the power meter for high duty cycles.

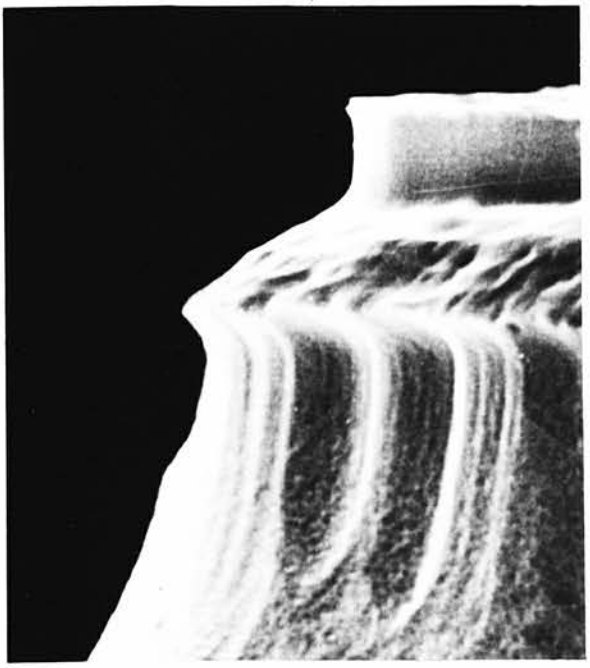
Although this equipment enabled satisfactory detection of microwave generation within the range 9-10 GHz, it would not readily permit a determination of maximum diode output. The original intention had been to establish a coupling coefficient for the cavity and a loading limit for oscillation, however these parameters depended on cavity length and the extensive calibration required to cover the entire operational range was considered too time consuming at a stage when the diode generation capability was still uncertain.

5.5 DIODE TEST RESULTS

Fabrication of matrix 6-11 was performed according to the culminating process schedule of Chapter 4, in the realisation of 5x5 array of stud mounted diodes.

Intermediate confirmation of correct alloy depth was obtained before heat sink attachment by mesa etching a portion of the silicon sample according to the process described in Section 4.1.5. The result of this test is illustrated in Plate 5.5.1, a scanning electron micrograph showing the desired p^+nn^+ structure with a physical p^+n junction at an alloy penetration of 2.3 μm into a total epitaxy thickness of 6 μm.

Insertion of a completed diode into the axial conductor



$P^+N\,N^+$ Mesa Profile

(x5400)

Plate 5.5.1

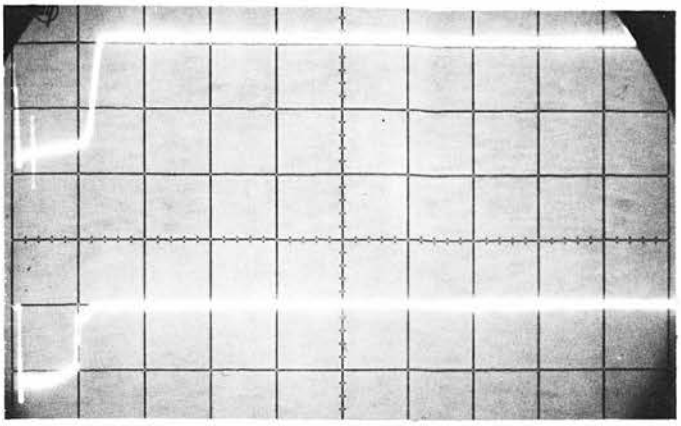
of the microwave cavity was followed by end plate fixture and advancement of the bias connecting screw to obtain circuit continuity, observed on a Tektronix curve tracer temporarily attached to confirm the suitability of d.c. characteristics before further testing.

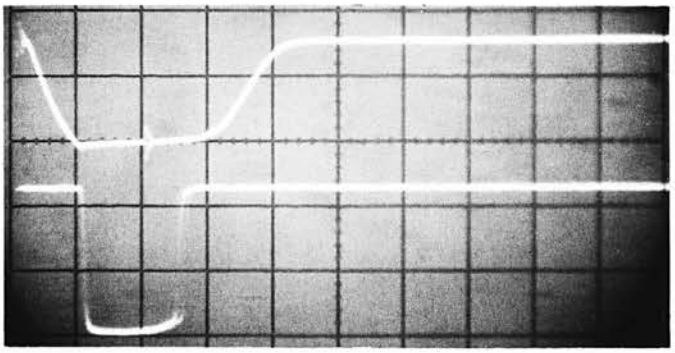
Within the matrix, several diodes exhibited excessive leakage in additon to premature breakdown, and these were retained for later diagnosis. Reverse bias characteristics comparable with development results were obtained however, showing an improved slope resistance attributable to the reduction in bulk n^- region. It was from devices exhibiting this performance that tests were made for microwave output, where the bias current was increased from 7 mA at a mark to space ratio of 1:10 and a 1 kHz pulse repetition frequency (p.r.f.).

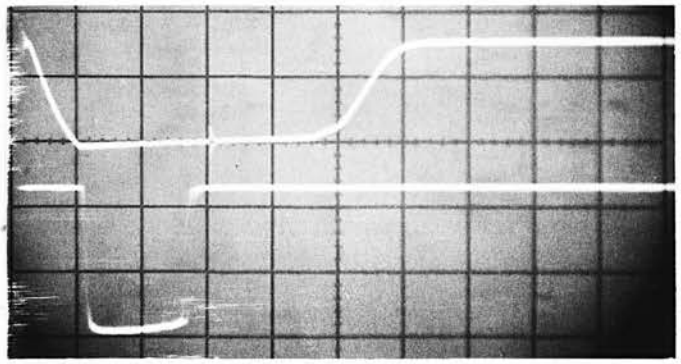
While loading was initially modest, X-band output could not be detected below a current threshold of 15 mA, and additional component failure occurred over this increment. Such devices were also subjected to later examination.

From the remaining diodes, microwave generation was finally detected within the range 8.8 - 10.2 GHz by suitable adjustment of biasing and cavity resonance. Limitations in output power and the restrictions in duty cycle can best be described by reference to Plates 5.5.2 and 5.5.3 composed of test results from two typical diodes.

Oscillograph 5.5.2a derived from a diode current pulse



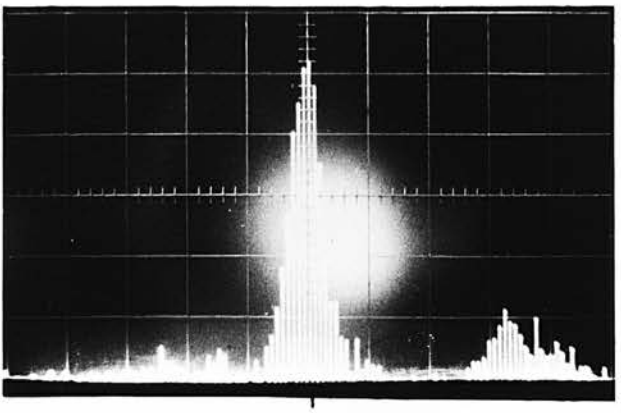




IMPATT Diode Oscillator Response

Plate 5.5.2

(a)



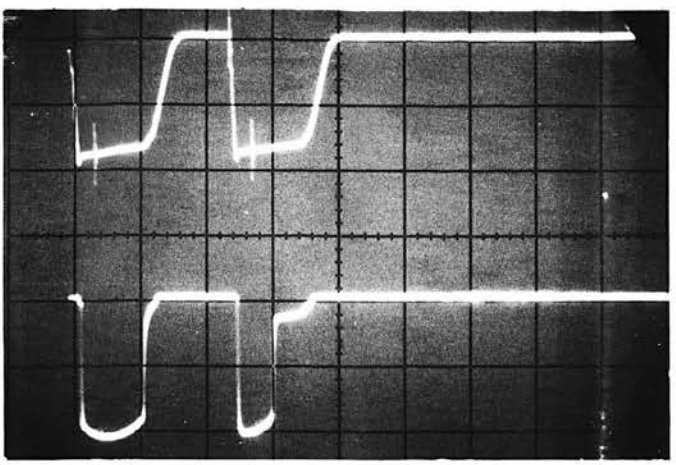
$f_0 = 9{\cdot}4$ GHz.

vert. deflection uncalibrated

horiz. dispersion = 10 kHz./div.

horiz. sweep rate = 0·1 div./ms.

(b)



horiz. scale = 0·1 ms./div.

IMPATT Diode Oscillator Response

Plate 5.5.3

of 15 mA shows an output of 3.9 mW detected within the -10 dB coupler, while the cavity was tuned to resonate at 9.1 GHz. Alternative cavity tuning revealed no comparable output level on swept spectrum analysis.

In order to confirm an apparent threshold for generation, the input pulse rise and fall times were increased, and oscillograph 5.5.2b demonstrates that output did notoccur below a current level of 15 mA.

Another feature emerged from this test performed with a wider pulse, namely the premature extinction of microwave output during the pulse interval. This limitation may be more clearly seen in oscillograph 5.5.2c indicating microwave generation for 0.17 ms during a 0.4 ms 15 mA pulse.

Tests performed on a different device to ascertain the maximum attainable generation level revealed that 36 mA pulsed bias current at a p.r.f. of 1 kHz gave rise to a total output of 135 mW which could be sustained for 0.1 ms. Spectrum analysis in the region of resonance at 9.4 GHz is displayed in Plate 5.5.3a. At this level of operation, additional output could be detected at several frequencies within the 9.1-9.6 GHz range, however the signal in every instance was considerably less than that present at the resonant frequency.

Final observation of the output extinction property was made in order to aid result appreciation. By applying current to

a diode in a double pulse mode at 1 kHz 'group' repetition rate, it was established that maximum output could be sustained from both pulses, each of the original 0.1 ms limiting duration, providing that a minimum separation of 0.1 ms was maintained. The effect of reduction in recovery period between pulses was to curtail the useful output as demonstrated in Plate 5.5.3b.

5.6 RESULT INTERPRETATION AND ASSOCIATED PRACTICAL DEVELOPMENT

5.6.1 D.C. DIODE CHARACTERISTICS

During development work on junction formation, the d.c. performance of alloyed regions was given serious consideration. Excessive reverse bias leakage was ascribed to surface effects, and non-uniform breakdown attributed to the well known microplasma phenomenon.

Because of the differences in physical structure between development samples and the final mesa diodes, these generalisations were made in preference to an early investigation of precise fault origins. Nevertheless, such an examination performed on finally tested diodes is worth description and imperative to future improvement of a fabrication schedule.

The mesa structure was inherently vulnerable to surface contamination since the junction periphery had been defined by chemical etching, and residual impurities removed by subsequent cleaning operations. Devices exhibiting excessive leakage were subjected to an additional mild etching cycle and a

rigorous cleaning treatment prior to further d.c. testing. In many instances an improved diode characteristic was observed and microwave generation could be achieved. Several components did not respond to this processing however, and some 'bulk' failure mechanism was obviously responsible.

It has been suggested by Chynoweth and Pearson[51] that the presence of silicon crystal dislocations within the space charge region will reduce the electron energy required for impact ionisation and hence the applied voltage necessary to initiate localised breakdown. Such dislocations might have been present within the parent substrate, created as defects in epitaxial growth, or initiated from mechanical damage introduced during slice attachment to the heat sink and thermocompression ball bonding of the gold wire top contact. Attempts to anneal out mechanical damage proved unsuccessful because of temperature restrictions associated with the presence of gold, a fast diffuser in silicon, and the likelihood of other process imperfections. The existence of defects within the parent substrate and epitaxial silicon may well have given rise to non-uniform alloy formation causing filamentary breakdown, while contaminants adsorbed on the silicon surface during atmospheric exposure prior to aluminium deposition and impurities within the evaporant might have resulted in recombination centres reducing carrier lifetime and increasing reverse saturation current.

Failures from unsatisfactory alloying or crystal damage were to some extent interdependent therefore, and could not be

readily distinguished in complete devices. These diodes were irredeemable and constituted a yield problem which could only be solved by closer process examination.

In addition to devices which possessed 'soft' reverse characteristics from surface contamination or multiple microplasma imperfections, several components demonstrated satisfactory low level d.c. performance but failed under a modest current increment. The effect of further chemical etching was to improve current handling capabilities of some but not all such diodes: accordingly, two distinct failure mechanisms were considered responsible.

The significance of diode edge profile has been discussed by Davies and Gentry[52] who indicated that mesa devices having a decreasing cross section from the heavily doped into the lightly doped junction region would exhibit preferential internal breakdown because of less intense edge fields. While the mesa configuration adopted during our work was initially expected to possess the desired geometry, it transpired that such a simple shape could not be realised and it was conceivable that some devices, having an adverse surface profile, would exhibit localised edge breakdown. Although no microscopic examination was made to substantiate this suggestion, the electrical improvements observed as a result of additional chemical etching indicated a modification in device contour rather than the removal of contamination associated with excessive leakage in other diodes.

Cases of premature d.c. failure which could not be rectified by chemical processing were attributable to localised internal breakdown which had caused excessive current densities and junction destruction. This condition had been recognised during early development of alloying, when different junction areas were found to exhibit similar current handling capabilities. Crystal damage introduced in recrystallisation was accordingly minimised by selection of an appropriate alloying cycle and resultant junction improvements would indicate that ultimate defects in device formation originated during later stages in the fabrication schedule. Although the nature of component failure was identical to that earlier described for devices exhibiting premature internal breakdown, satisfactory low level reverse bias characteristics suggested the presence of a single microplasma which was responsible for filamentary breakdown.

5.6.2 MICROWAVE GENERATION - LIMITATIONS AND ATTEMPTED IMPROVEMENTS

Detection of microwave output within the range 8.8-10.2 GHz demonstrated the feasibility of alloying and integral slice attachment as process stages in IMPATT diode manufacture. Because of severe limitations in generated output however, it was necessary to confirm that restrictions were imposed by inadequate device heat sinking, and not by unsatisfactory junction properties.

The construction of laboratory samples had not included an encapsulation stage, and electrical connection to each device

cathode was achieved by means of pressure contact on a gold wire which had been thermocompression ball bonded to a deposited gold electrode. Within the test equipment therefore, thermal conduction from a device junction occurred mainly through the screwed copper mount to the adjacent microwave cavity assembly: diodes were attached directly to a steel central conductor of low mass and thermal conductivity as a result of which it was anticipated that maximum device dissipation would be extremely modest.

Since current pulses in excess of 36 mA and 100 μs duration caused extinction of X-band output, an examination was made of a d.c. diode characteristic at this current level. Measurements obtained from a Tektronix curve tracer revealed a breakdown voltage of 100V at 36 mA (compared with 80V at low current levels) indicating an excessive rise in junction temperature. This was confirmed by mounting the diode in a large copper heat sink outwith the cavity, and performing the same test when a reverse current of 36 mA could be sustained with 85V applied.

In an attempt to increase the available X-band output, the central steel conductor in the microwave cavity was replaced with a copper rod. Although the permissible current level rose by about 20%, the coupled output showed a less appreciable increase, presumably since the resonant mode energy distribution had changed to favour TEM excitation.

While the global nature of this project was under initial discussion, it appeared appropriate to terminate the work by constructing a microstrip oscillator which would demonstrate IMPATT diode operation in a microelectronic microwave medium. As the time allocation for experimental work had already been exceeded, a final endeavour to solve the heat sinking problem was related to this desire to utilise microstrip, in the knowledge that devices mounted in a substantial copper base attached to the ground plane would be capable of the necessary dissipation.

It was proposed that a dielectric substrate, metallised on both surfaces, should be subjected to photolithographic and chemical etching operations in order to define ground plane and stripline configurations. Ultrasonically drilled holes would then allow microstrip attachment to a copper base and provide a feed-through for the diode.

Unfortunately, the 95% alumina material available for this work was not immediately suitable for accurate pattern delineation, being slip cast and having a surface undulation of $\pm$ 125 μm over the 3 cm. x 2 cm. substrate area. It was found that stock removal by sequential lapping with 12 μm and 3 μm diamond compounds on a cast iron plate reduced this irregularity to less than $\pm$ 20 μm , but gave rise to a further problem of substrate pitting where certain crystalline regions had been 'plucked' from the alumina. Although attempts to eliminate the presence of such voids (measuring about 50 μm in diameter and 15 μm in depth)

proved unsuccessful, gentle and lengthy abrasion with 3 μm diamond impregnated microcloth provided a substrate possessing a surface flatness better than ± 30 μm , surface finish better than ± 0.15 μm and an incidence of voids less than 30 μm in diameter and 10 μm in depth.

Since it was unlikely that the alumina material could be further improved superficially, an attempt was made to establish the feasibility of microstrip using gold metallisation which would ultimately allow thermo-compression attachment of a diode contact wire. The initial procedure involved

(a) vacuum deposition of a nichrome thin film to improve substrate adhesion properties (film thickness ≃ 200Å , substrate temperature = 150°C)

(b) vacuum deposition of gold to a film thickness >1.5 μm by multiple source evaporation (substrate temperature = 150°C)

(c) photolithographic definition of a conductor pattern including line widths of 50 μm which would be required for high r.f. impedance, d.c. bias circuitry.

(d) chemical removal of unwanted gold and nichrome.

During the protracted etching period required to dissolve

such a thickness of gold, photoresist degradation and under-etching resulted in complete removal of all lines less than 125 μm wide. Consequently, the schedule was modified to reduce gold deposition to 1500 Å , in the hope that accurate pattern definition would be possible and that an electroplating procedure could be used to establish the requisite conductor thickness. Under these circumstances however, the existence of voids within the alumina surface gave rise to discontinuities in narrow conductors and work could not proceed to the electroplating stage.

Clearly, the substrate material was still inadequate despite considerable efforts to improve its surface properties. Further pursuit of this lengthy, fundamental, and essentially preliminary investigation was unlikely to enable microstrip formation in the short term, if at all; and since the acquisition of superior substrate material was to involve an unacceptable delay, it was decided that practical development would have to be terminated.

5.7 CONCLUDING REVIEW AND RECOMMENDATIONS FOR FUTURE WORK

During a preliminary literature survey on the IMPATT diode, it was realised that considerable engineering effort had already been expended on a device possessing immediate commercial significance, and that the natural progression from research into development and production media within a short space of time would inevitably result in an escalation of manpower engaged in diode manufacture. For these reasons, and with the limited

resources at our disposal, it seemed unlikely that a meaningful contribution could be made along similar lines: accordingly in this study, emphasis was placed on abrupt junction formation and the development of fabrication techniques which might achieve greater status within the semiconductor industry.

In an examination of low temperature silicon doping, the diverse technologies of double epitaxy and ion implantation were considered suitable vehicles for IMPATT diode production, being recently conceived techniques and having far reaching potential in other spheres of semiconductor device manufacture. It was unfortunate however, that the impracticality of each in turn did not emerge until late in the development of experimental systems: our departmental epitaxial reactor became committed to the completion of other research, and the financial demands of an implantation system proved too great.

By contrast, the essential equipment required for aluminium silicon alloy doping was extremely modest and readily available. Because this process had been used in earliest work on solid state diode fabrication and had become almost obselete in its infancy with the development of a high temperature diffusion technique which offered vastly increased versatility, it was necessary to reconstruct a controlled alloying facility in our laboratory which would permit consistently uniform junction formation giving predictable silicon penetration. Although this was achieved for small sample areas, severe restrictions on

ultimate device yield were imposed by an axial temperature gradient within the furnace tube and the limitations on accuracy of thermal cycle reproduction. Further development of alloy junction formation would involve greater furnace sophistication incorporating independent proportional temperature control of two or three heating elements enclosing the reaction zone, and a mechanised sample conveyor belt to enable a reproducible range of thermal cycles by programmed variation in transport velocity.

Measurement of alloy penetration was achieved by conventional lap and stain techniques, and by direct measurement on mesa structures. The latter method was made possible by scanning electron microscopy which revealed contour transitions between alloyed, n^-, and n^+ regions; in addition to permitting detailed scrutiny of surface damage during an assessment of chemical etching behaviour. This facility proved extremely useful and would enable a future contribution to the work by Calverley et al.[43] on the dependence of mesa profile on etchant constituents.

Methods of diode isolation and separation were considered in conjunction with ultimate device requirements of high junction dissipation and test circuit compatibility, while a desire to achieve low cost manufacture was given high priority. The concepts of extending integral slice fabrication to include a mounting/heatsinking stage and employing ultrasonic drilling in device definition and separation appeared attractive in eliminating individual device assembly and the need for a sophisticated

photolithographic facility. Although careful consideration was given to the work by Swan[48] on improved IMPATT diode performance from devices mounted on type *II*A diamond, a similar approach was rejected in favour of employing more practical manufacturing materials.

Fabrication of gold plated copper heatsinks enabled the development of integral slice thermocompression bonding according to the technique outlined by Hambleton. Although little difficulty was experienced in obtaining satisfactory component adhesion, the convenient extravagance of a precision press during experimental stages could not be justified for future work: it remains therefore to ascertain the true complexity required to achieve acceptable results before an accurate impression of process costs can be realised.

In contrast to the early success with thermocompression slice bonding, initial endeavours to use ultrasonic drilling for device definition and separation revealed a high incidence of sample disintegration, consistently irregular mesa surfaces, and an inability to machine the malleable heatsink material. While it became obvious that ultrasonic abrasion of copper was impractical, the desire to eliminate photolithography and chemical stock removal from device fabrication prompted a closer examination of silicon machining properties on the assumption that an alternative method of heatsink division could be developed. Unfortunately, the presence of copper beneath the silicon sample had an adverse effect on erosion rate, and the required mesa

geometry could only be obtained with coarse abrasive which produced excessive surface damage.

As an attempted compromise, preliminary ultrasonic drilling was used to define devices, and subsequent chemical etching applied to complete the separation process. Because of detrimental lateral etching, resultant geometries were unacceptable and the involvement of ultrasonics had to be terminated. Our experiences had indicated the feasibility of precision silicon machining; and its potential usefulness in conjunction with less malleable mounting materials might be realised in later work. It must be mentioned that a degree of silicon surface damage is inevitable with this process, however the use of fine abrasives would necessitate only slight remedial polishing.

As an expedient to define a diode matrix and achieve device separation, selective chemical etching was latterly employed on a thinned silicon slice which had been prepared with a masking array of gold wire thermocompression ball bonds. Although this measure was considered acceptable under the pressing circumstances of a protracted development program, future device manufacture would make use of photolithographic masking to ensure consistent diode junction areas and a regular component array.

Heatsink division was accomplished using a milling procedure designed to minimise cutting waste, and individual

components were manually threaded to provide a means of circuit insertion. Since the threading operation proved extremely tedious and a serious threat to high component yield, it would be necessary to consider alternative methods of circuit attachment at a time during future development when the interchangeability of diodes became less important.

Dynamic device testing was performed in a right circular cylindrical microwave cavity with d.c. bias applied to the diode from a constant current source designed for pulse or continuous operation. Although the sidewall cavity coupling configuration was adequate for microwave detection purposes, it did not readily permit a comparison of diode parameters, since the coupling coefficient was a function of the variable cavity length. Improvements could be expected however, on changing to iris/ waveguide coupling from the fixed end plate, provided that adequate precaution is taken to maintain d.c. isolation.

Severe limitations in diode power dissipation arose because of unsatisfactory mounting within the cavity: this was inferred from the observed increase in current handling capabilities of diodes attached to a large copper heatsink. In retrospect, the unproductive time spent attempting to incorporate such heatsinking in a microstrip oscillator would have been better directed towards the forementioned improvements in cavity coupling; and at the expense of d.c. supply modifications, the transfer of diode mounting from the coaxial conductor to the end

plate. Furthermore, the possibility of dominant TEM excitation mentioned in Section 5.6.2 should be investigated more fully, since restricted oscillator output might also be attributable to minimal coupling to this alternative operating mode. By reorientating the output loop to increase coupling to an angular magnetic field component, it would be possible to establish the validity of this suggestion, and perhaps achieve vastly improved oscillator performance.

In conclusion, an attempt has been made to introduce a novel batch manufacturing technique for abrupt junction silicon IMPATT diode fabrication. While alloy junction formation, integral slice thermocompression mounting, and device formation have all been successfully accomplished, it should be noted that some of the development problems and their devious solutions have altered in significance as a result of recent technological advances. For example, the present availability of extremely thin silicon slices (< 50μm) eliminates the need for additional stock removal and increases the credibility of chemical mesa formation as opposed to further development of ultrasonic drilling. Nevertheless, the work undertaken in this project demonstrates the usefulness and limitations of several low cost manufacturing techniques, which, with the addition of a suitable encapsulation procedure, should enable production of high quality IMPATT diodes for a modest capital investment.

ACKNOWLEDGEMENTS

I wish to thank my supervisors, Professor W.E.J. Farvis and Dr. R.D. Pringle for their direction and encouragement during this work.

The assistance of Ferranti Limited in financing the study and providing clerical assistance has been greatly appreciated.

APPENDIX A

Derivation of Depletion Length in a Reverse Biased Abrupt Junction

Under reverse bias, the one sided abrupt junction represented in Fig. A.1 will exhibit carrier depletion in both directions from the physical interface between impurity types. Because of the higher p-type doping level, depletion will predominate within the n-region and a charge continuity equation may be written:

$$qN_a A\,dp = qN_d A\,dn \qquad \text{A.1}$$

where N_a and N_d are the acceptor and donor carrier concentrations, dn and dp are the depletion lengths and A is the junction area.

The one dimensional Poisson equation may be written:

(a) for negative x ,

$$\frac{d^2V}{dx^2} = \frac{qN_a}{\varepsilon} \qquad \text{A.2a}$$

(b) for positive x ,

$$\frac{d^2V}{dx^2} = -\frac{qN_d}{\varepsilon} \qquad \text{A.2b}$$

Integration of A.2a, b with the boundary conditions that

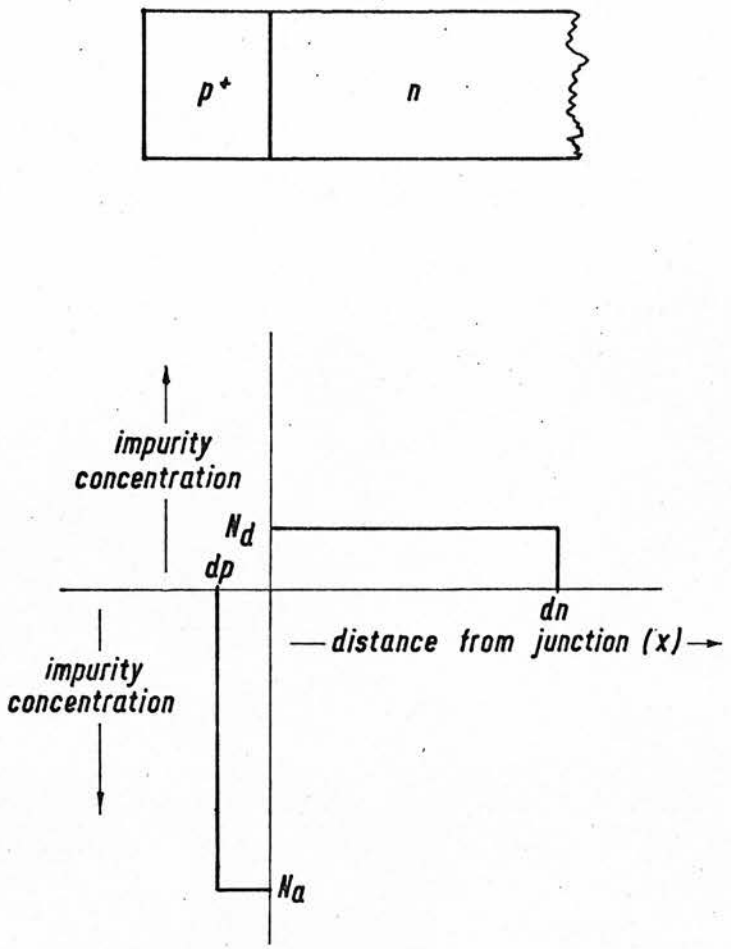

p+
n
impurity
concentration
Nd
dp
dn
distance from junction (x)
impurity
concentration
Na

Fig. A.1

electric field strength is zero at $x = dp$ and $x = dn$ gives:

(a) for negative x,

$$\frac{dV}{dx} = \frac{qN_a}{\varepsilon}(x - dp) \qquad \text{A.3.a}$$

(b) for positive x,

$$\frac{dV}{dx} = -\frac{qN_d}{\varepsilon}(x - dn) \qquad \text{A.3.b}$$

Further integration yields equations for voltages

$$x \leq 0, \qquad V_p = \frac{qN_a}{2\varepsilon}\left(\frac{x^2}{2} - dp.x\right) + C_p \qquad \text{A.4.a}$$

$$x \geq 0, \qquad V_n = -\frac{qN_d}{2\varepsilon}\left(\frac{x^2}{2} - dn.x\right) + C_n \qquad \text{A.4.b}$$

Since at $x = 0$ equation 3 a,b must be equal, the integration constants must equate. The voltages at $x = dp$ and $x = dn$ may therefore be expressed as

$$V_{dp} = -\frac{qN_a}{2\varepsilon} dp^2 + C \qquad \text{A.5.a}$$

$$V_{dn} = \frac{qN_d}{2\varepsilon} dn^2 + C \qquad \text{A.5.b}$$

and the total voltage V_t across the junction is

$$V_t = V_{dn} - V_{dp} = \frac{q}{\varepsilon}(N_d\, dn^2 + N_a dp^2) \qquad \text{A.6}$$

Substitution from equation 1.1 gives

$$dn = \sqrt{\frac{2\varepsilon V_t}{q} \cdot \frac{N_a}{N_a N_d + N_d^2}} \qquad \text{A.7.a}$$

$$dp = \sqrt{\frac{2\varepsilon V_t}{q} \cdot \frac{N_d}{N_a N_d + N_a^2}} \qquad \text{A.7.b}$$

and hence the total depletion width (W)

$$W = dp + dn = \sqrt{\frac{2\varepsilon V_t}{q(N_d + N_a)}}\left[\sqrt{\frac{N_a}{N_d}} + \sqrt{\frac{N_d}{N_a}}\right] \qquad \text{A.8}$$

This general equation can be reduced for the one sided abrupt junction where $N_a \gg N_d$ to

$$W = \sqrt{\frac{2\varepsilon V_t}{qN_d}} \qquad \text{A.9}$$

APPENDIX B

Derivation of the Ionisation Integrals under Avalanche Conditions

Acceleration of electrons and holes in a high electric field can cause electron-hole pair generation by collision with electrons in the valence band of the semiconductor. The rate of ionisation is defined as $\alpha(E)$ being the average number of ionising events per centimetre travel in the field: thus an electron will create N electron-hole pairs in its travel across a space charge width W where

$$N = \int_0^W \alpha(E)\, dx$$

If it is assumed for the initial analysis that $\alpha(E)$ has the same value for electrons and holes, then N electron-hole pairs can generate N^2 etc.

In this way a reverse saturation current I_o will be multiplied to a total current I where

$$I = MI_o = I_o(1 + N + N^2 \cdots) = \frac{I_o}{1 - N} \qquad (N < 1) \qquad \text{B.1}$$

and $M = \frac{1}{1 - N}$ is called the multiplication factor.

Breakdown will occur where M becomes infinite i.e.

$$N = \int_0^W \alpha(E) = 1 \qquad \text{B.2}$$

Consideration is now given to Fig.B.1 which demonstrates current flow through a p-n junction region. A hole current entering the junction at $x = 0$ multiplies to become I_p at x and $I_p + dp$ at $x + dx$. Electrons generated between $x + dx$ and W give rise to an electron current of I_n at $x + dx$, and $I_n + dI_n$ at x. Consequently an increase in hole current over the distance dx can be expressed as

$$dI_p = \alpha_p I_p \, dx + \alpha_n I_n dx \qquad \text{B.3}$$

$$\text{or} \qquad \frac{dI_p}{dx} = \alpha_n I + (\alpha_p - \alpha_n) I_p \qquad \text{B.4}$$

where α_n, α_p are the individual ionisation rates for electrons and holes, and $I = I_p + I_n$

Substitution of boundary conditions $I_p(0) = I_{ps}$ and $I_p(W) \simeq I$ together with the standard solution for equations of the form $y' + Py = Q$, enables equation B.4 to be integrated:

$$I_p = I \frac{\left[\frac{1}{M_p} + \int_0^x \alpha_n e^{-\int_0^x (\alpha_p - \alpha_n) dx'} dx \right]}{e^{-\int_0^x (\alpha_p - \alpha_n) dx'}} \qquad \text{B.5}$$

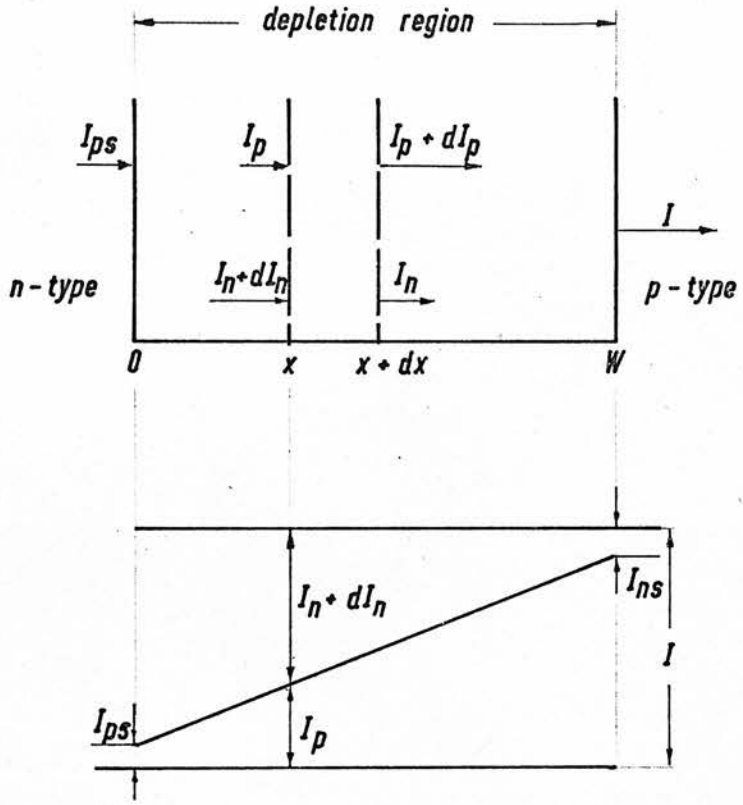

depletion region
I_{ps}
I_p
$I_p + dI_p$
I
n - type
$I_n + dI_n$
I_n
p - type
0
x
x + dx
W
$I_n + dI_n$
I_{ns}
I
I_{ps}
I_p

Fig. B.1

where M_p is the multiplication factor of holes, defined as

$$M_p = \frac{I_p(W)}{I_p(0)}$$

Equation B.5 may be written as

$$1 - \frac{1}{M_p} = \int_0^W \alpha_p \, e^{-\int_0^x (\alpha_p - \alpha_n) dx'} dx \qquad \text{B.6}$$

Since the avalanche voltage corresponds to the multiplication factor approaching infinity, the ionisation integral at breakdown is given by

$$\int_0^W \alpha_p \, e^{-\int_0^x (\alpha_p - \alpha_n) dx'} dx = 1 \qquad \text{B.7}$$

In the case of electron initiated avalanche, the appropriate equation is

$$\int_0^W \alpha_n \, e^{-\int_0^x (\alpha_n - \alpha_p) dx'} dx = 1 \qquad \text{B.8}$$

which may be derived similarly.

APPENDIX C

The Determination of Impurity Profile from Diffusion into Epitaxial Silicon

During the processes of n-type epitaxial growth and p^+ diffusion, out-diffusion of impurity atoms from the substrate material will take place and the resultant doping profile can be assessed from the theory presented by Rice.

The analysis will be considered in two stages:

(1) Epitaxial growth upon an antimony doped silicon slice (10^{19} atoms cm^{-3}) to a thickness of 5.7 μm with an arsenic impurity concentration of 6.10^{15} atoms. cm^{-3}.

(2) Diffusion of boron from an infinite impurity source (5.10^{20} atoms. cm^{-3}) to form an electrical junction with the epitaxial layer at a distance 2.0 μm from the silicon surface.

Epitaxial Growth

At a temperature of 1200°C, the $Si\ Cl_4$ deposition process enables epitaxial silicon growth at a rate of typically 0.7 μm min^{-1}. This implies a deposition time of 6 minutes from which the following tabulated out-diffusion values can be calculated.

Antimony Impurity Concentration (Atoms cm^{-3})	Distance from Epitaxy Surface (microns)
10^{15}	4.90
10^{16}	5.06
$5.0\ 10^{18}$	5.70
10^{19}	6.85

A similar analysis of arsenic in-diffusion gives

Arsenic Impurity Concentration (Atoms cm^{-3})	Distance from Epitaxy Surface (microns)
$6.0.\ 10^{15}$	5.22
$4.5.\ 10^{15}$	5.60
$3.0.\ 10^{15}$	5.70
$6.0.\ 10^{12}$	6.00

Diffusion

From equation 3.1.2.1, the process time required to achieve a p-n junction in the epitaxy by boron diffusion at $1200^{o}C$ from an infinite impurity source of 5.10^{20} cm^{-3} is 8 minutes. A complementary error function profile can be plotted

Boron Impurity Concentration (atoms cm^{-3})	Distance from Epitaxy Surface (microns)
$5.0.\ 10^{20}$	0
10^{20}	0.5
10^{19}	1.0
10^{18}	1.4
$6.0.\ 10^{15}$	2.0

In order to approximate the effect on epitaxy and substrate profiles in the vicinity of the physical interface, it is necessary to consider the effect of a further 8 minutes diffusion at 1200°C, i.e. an equivalent total epitaxial deposition period of 12 minutes.

The modified impurity concentrations can now be estimated.

Antimony Impurity Concentration (atoms cm^{-3})	Distance from Epitaxy Surface (microns)
10^{19}	7.4
$5.0.\ 10^{18}$	5.7
10^{16}	4.7
10^{15}	4.4

Arsenic Impurity Concentration (atoms cm^{-3})	Distance from Epitaxy Surface (microns)
$6.0.\ 10^{15}$	5.0
$3.0.\ 10^{15}$	5.7
$6.0.\ 10^{12}$	6.3

APPENDIX D

Silicon Slice Preparation I

1. Immerse in warm (fuming) concentrated sulphuric acid for 10 min.

2. Transfer to boiling double-distilled, deionised water for 5 min.

3. Transfer to boiling concentrated nitric acid for 10 min.

4. Transfer to boiling double-distilled, deionised water for 5 min.

5. Transfer to a room temperature solution of 1 part hydrofluoric acid: 5 parts double-distilled, deionised water for 30 s.

6. Rinse in recirculating double-distilled, deionised water for 15 min.

7. Dry in a nitrogen atmosphere.

APPENDIX E

Silicon Slice Preparation II

1. Immerse in a warm solution of 1 part formic acid: 1 part hydrogen peroxide for 15 min.

2. Rinse in recirculating double distilled, deionised water for 2 min.

3. Transfer to a solution of 1 part hydrochloric acid: 1 part hydrogen peroxide: 1 part water for 5 min.

4. Rinse in recirculating double-distilled, deionised water for 10 min.

5. Transfer to a warm solution of 1 part hydrogen peroxide: 1 part ammonium hydroxide: 5 parts water for 10 min.

6. Rinse in recirculating double distilled, deionised water for 2 min.

7. Transfer to a room temperature solution of 1 part hydrofluoric acid: 10 parts water for 30s.

8. Rinse in recirculating doube distilled, deionised water for 15 min.

9. Dry in a nitrogen atmosphere.

APPENDIX F

Operation Sequence for Vacuum Deposition of Aluminium

1. Evacuate vacuum chamber to $< 5.10^{-5}$ torr.
2. Outgas sample and holder at 250°C for 15 min.
3. Reduce sample temperature to 100°C and allow system pressure to reduce to $< 5.10^{-5}$ torr.
4. Commence liquid nitrogen feed to cold trap.
5. Allow system pressure to become $< 10^{-5}$ torr.
6. Apply source heating until aluminium evaporation commences.
7. Remove protective shutter from vapour stream.
8. Continue evaporation to source dryness.
9. Reduce source temperature.
10. Reduce sample temperature to < 50°C.
11. Discontinue liquid nitrogen supply to cold trap.
12. Return vacuum chamber to atmospheric pressure.

REFERENCES

1. WHEELER, H.A. — I.E.E.E. Trans. Microwave Theory and Techniques MTT-13 (1965) p.172

2. YAMASHITA, E. & MITTRA, R. — I.E.E.E. Trans. Microwave Theory and Techniques MTT-16 (1968) p.251

3. PRESSER, A. — Microwaves (March 1968) p.55

4. SHOCKLEY, W. — Bell Syst. Tech. J. 33 (1954) p. 799

5. READ Jr., W.T. — Bell Syst. Tech. J. 37 (1958) p. 401

6. JOHNSTON, R.L. et. al. — Bell Syst. Tech. J. 44 (1965) p.369.

7. LEE, C.A. et. al. — Appl. Phys. Lett. 6 (1965) p.89

8. MISAWA, T. — I.E.E.E. Trans. Electron Devices ED-13 (1966) p.137

9. MISAWA, T. — I.E.E.E. Trans. Electron Devices ED-13 (1966) p.143

10. GILDEN, M & HINES, M.E. — I.E.E.E. Trans. Electron Devices ED-13 (1966) p.169

11. KOVEL, S.R. & GIBBONS, G. — Proc. I.E.E.E. 55 (1967) p. 2066

12. SZE, S.M. & GIBBONS, G. — Appl. Phys. Lett. 8 (1966) p.111

13. SZE, S.M. — "Physics of Semiconductor Devices" Wiley-Interscience 1969.

14. THEURER, H.C. — J. Electrochem. Soc. 108 (1961) p.649

15. CAVE, E.F. & CZORNY, B.R. — R.C.A. Rev. 24 (1963) p.523

16. GROVE, A.S. et. al. — J. Appl. Phys. 36 (1965) p.802

17. KAHNG, D. et. al. — J. Electrochem. Soc. 109 (1962) p. 1106

18. RICE, W. — Proc. I.E.E.E. 52 (1964) p.284

19. GROSSMAN, J.J. — J. Electrochem. Soc. 110 (1963) p. 1065

20. GUPTA, D.C. & YEE, R. — J. Electrochem. Soc. 116 (1969) p. 1561

21. WILLIAMS, E.L. — J. Electrochem. Soc. 108 (1961) p.795

22. THURSTON, M.C. et. al. — Research Report No. AD 274969, Ohio State University

23. BEAN, K.E. & GLEIM, P.S. — Proc. I.E.E.E. 57 (1969) p.1469

24. WOLF, H.E. — "Silicon Semiconductor Data", International Series of Monographs in Semiconductors 9 (1969) Pergamon Press.

25. JOYCE, B.A. & BRADLEY, R.R. — J. Electrochem. Soc. 110 (1963) p. 1235

26. BHOLA, S.R. & MAYER, A. — R.C.A. Review 24 (1963) p.512

27. RICHMAN, D. and ARLETT, R.H. — J. Electrochem. Soc. 116 (1969) p.872

28. RICHMAN, D. et. al. — R.C.A. Review 31 (1970) p.613

29. STOVER, H.L. — Electronics 41 (1968) p.96

30. SHOCKLEY, W. "Forming Semiconductive Devices by Ionic Bombardment" U.S. Patent 2787564 (1954).

31. CUSSINS, W.D. Proc. Phys. Soc. 68 (1955) p.213

32. ALVAGER, T. & HANSEN, N.J. Rev. Sci. Instr. 33 (1962) p. 567

33. WILSON, R.G. Record I.E.E.E. 9th Annual Symposium on Electron, Ion, and Laser Beam Technology. Berkeley (1967) p.22

34. THONEMAN, P. et. al. Proc. Phys. Soc. (London) 61 (1948) p.483

35. MOAK, C.D. et. al. Nucleonics 9, 3, (1951) p.18

36. GIBBONS, J.F. Proc. I.E.E.E. 56 (1968) p.295

37. FERBER, R.R. I.E.E.E. Trans. Nucl. Sci. NS-10 (1963) p.15.

38. MARTIN, F.W. et. al. I.E.E.E. Trans. Nucl. Sci. NS-11 (1964) p.280

39. KLEINFELDER, W.J. et. al — Record. of I.E.E.E. 9th Annual Symposium on Electron, Ion and Laser Beam Technology (1967) p.101

40. PEARSON, G.L. & SAWYER, B. — Proc. I.R.E. 40 (1952) p. 1348

41. BURGER, R.M. & DONOVAN, R.P. — "Fundamentals of Silicon Integrated Device Technology", 1, Prentice-Hall 1967

42. NICOLAU, I.F. — Solid State Electronics 12 (1969) p.446

43. CALVERLEY, A. et. Al. — S.E.R.L. Tech. J. 20 3, (1970) p.11.1

44. LINDMAYER, J. & WRIGLEY, C.Y. — "Fundamentals of Semiconductor Devices" D. Van Nostrand Coy. Inc. 1965.

45. MATLOW, S.L. & RALPH, E.L. — J. Appl. Phys. 30 (1959) p. 541

46. KRESSEL, H. — R.C.A. Review 28 (1967) p. 175

47. HAMBLETON, K.G. "The Fabrication of Semiconductor Diodes on Copper Heat Sinks". Presentation at 3rd Conference on Solid State Devices, Exeter 1969.

48. SWAN, C.B. Proc. I.E.E.E. 55 (1967) p. 1617

49. MONTGOMERY, C.G. "Technique of Microwave Measurements" Radiation Laboratory Series, II, McGraw-Hill Book Company Inc. 1947.

50. MORTON, I.F. "The Microwave Frequency Conductivity of Thin Metal Films". M.Sc. Thesis, University of Edinburgh 1969.

51. CHYNOWETH, A.G. & PEARSON, G.L. J. Appl. Phys. 29 (1958) p. 1103

52. DAVIES, R.L. & GENTRY, F.E. I.E.E.E. Trans. Electron Devices ED-11 (1964) p.313

53. GUMMEL, D.L. & SCHARFETTER, H.K. — I.E.E.E. Trans. Electron Devices ED-16 (1969) p.64

54. KUTATELADZE, S.S. — "Fundamentals of Heat Transfer" Edward Arnold (Publishers) Ltd. (1963)

55. DECKER, D.R. & SCHORR, A.J. — I.E.E.E. Trans. Electron Devices ED-17 (1970) p.739